

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: December 4, 2001, 13:12:37 ; Search time 4574.73 Seconds  
(without alignments)  
1190.031 Million cell updates/sec

Title: US-09-508-821B-2

Perfect score: 330  
Sequence: 1 tccctcccccacccctccac.....ggccagggctactgcagcc 330Scoring table: IDENTITY\_NUC  
Gapop 10.0 ; Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

## Database :

GenEmbl: \*  
1: gb\_da: \*  
2: gb\_htg: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sto: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vl: \*  
15: em\_da: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_om: \*  
20: em\_or: \*  
21: em\_ov: \*  
22: em\_pat: \*  
23: em\_ph: \*  
24: em\_pl: \*  
25: em\_ro: \*  
26: em\_sts: \*  
27: em\_sy: \*  
28: em\_un: \*  
29: em\_vl: \*  
30: em\_htgo\_hum: \*  
31: em\_htgo\_inv: \*  
32: em\_htgo\_rod: \*  
33: em\_htg\_hum: \*  
34: em\_htg\_inv: \*  
35: em\_htg\_rod: \*  
36: em\_htg\_other: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result	Score	Match	Length	DB	ID	Description
1	329	99.7	330	6	AX004677	AX004677 Sequence
2	320.6	97.2	542	11	G09710	G09710 human SRS C
3	317	96.1	5915	9	AB058723	AB058723 Homo sapi
4	317	96.1	6022	6	AX004680	AX004680 Sequence
5	317	96.1	160544	2	HS253P07	HS253P07
6	306	92.7	23843	9	HS2471791	HS2471791 Homo sapi
7	283.6	85.9	5395	6	AX024201	AX024201 Sequence
8	283.6	85.9	5667	9	HS2471790	HS2471790 Homo sapi
9	288.6	78.4	595	11	AF021115	AF021115 Homo sapi
10	170.2	51.6	328	6	AX004676	AX004676 Sequence
11	170.2	51.6	7222	10	W05MA	W05MA
12	68.8	20.8	197232	2	AC073946	AC073946 Mus muscu
13	68.8	20.8	480	3	AF104706	AF104706 Plasmodiu
14	67.8	20.5	1968	10	MMU06463	MMU06463 Mus muscu
15	67.2	20.4	5120	6	AX067148	AX067148 Sequence
16	67.2	20.4	5120	9	AF152102	AF152102 Homo sapi
17	67.2	20.3	461	3	AF104705	AF104705 Plasmodiu
18	65.8	19.9	71803	2	AC010322	AC010322 Homo sapi
19	65.4	19.8	1547	10	MMU070649	MMU070649 Mus muscu
20	65.2	19.8	647	3	U91677	U91677 Plasmodium
21	65.2	19.8	1512	10	MMU070655	MMU070655 Mus muscu
22	65.2	19.8	1512	10	MMU070656	MMU070656 Mus muscu
23	65.2	19.8	1546	10	MMU070645	MMU070645 Mus muscu
24	65.2	19.8	1549	10	MMU070657	MMU070657 Mus muscu
25	65.2	19.8	1586	10	MMU070653	MMU070653 Mus muscu
26	65.2	19.8	1598	10	MMU070654	MMU070654 Mus muscu
27	65.2	19.8	10266	22	E11536	E11536 DNA from Y
28	65.2	19.8	14625	10	MSR10C	MSR10C Mus muscu
29	65	19.7	615	3	PFAM5A2L	PFAM5A2L
30	64.4	19.5	162869	9	AC090645	AC090645 Homo sapi
31	64.2	19.5	1611	10	MMU070651	MMU070651 Mus muscu
32	64.2	19.5	1612	10	MMU070652	MMU070652 Mus muscu
33	64	19.4	185306	2	AC013570	AC013570 Homo sapi
34	63.4	19.2	413	3	AF104711	AF104711 Plasmodiu
35	63.4	19.2	471	3	AF104712	AF104712 Plasmodiu
36	62.8	19.0	5719	10	MMU070651	MMU070651 Mus muscu
37	62.4	18.9	527	3	AF104713	AF104713 Plasmodiu
38	62.4	18.9	8071	3	MSOR1RRT	MSOR1RRT
39	62.2	18.8	460	3	AF329560	AF329560 Plasmodiu
40	61.8	18.7	503	3	AF104707	AF104707 Plasmodiu
41	61.6	18.7	1546	10	MMU070642	MMU070642 Mus muscu
42	61.6	18.7	1550	10	MMU070641	MMU070641 Mus muscu
43	61.6	18.7	1552	10	MMU070650	MMU070650 Mus muscu
44	61.6	18.7	1558	10	MMU070646	MMU070646 Mus muscu
45	61.6	18.7	2298	10	AF337043	AF337043 Mus muscu

## ALIGNMENTS

RESULT 1  
AX004677 330 bp DNA PAT 24-AUG-2000  
LOCUS Sequence 2 from Patent WO9915639.  
DEFINITION AX004677  
ACCESSION AX004677.1 GI:9928113  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
unidentified.  
unclassified.  
REFERENCE 1 (bases 1 to 330)  
AUTHORS Rouleau G.A. and Joober R.  
TITLE Polymorphic cag repeat-containing gene and uses thereof  
JOURNAL Patent: WO 9915639-A 2 01 APR 1999;  
ROUTEAU GUY A (CA); UNIV MCGILL (CA)  
FEATURES  
SOURCE  
1..330  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 72 a 128 c 83 g 46 t 1 others















```
/protein_id="raa06184.1"
/db_xref="GI:475016"
/translation="MQSFRGCGFHGKQNPQSOETSELYNQPCQAGLSCDROR
LLAKDYSPQPTVEGTTGGTGTGATVAADKTHRSKGLAGFAFFSYVQSSPFP
GRYSGEGLQWGGPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPP
RTSLHVPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPP
VGGGGAHSYKSCYTAQPHDRMANANLAPQPPQPPQPPQPPQPPQPPQPPQ
QGHHTQTHYONLAKYQVHGOOGVCPDPTAVRPEYQYTFSPSSPARSVG
RSPSYSPSPMLNENFYSPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ
QAGNCPQLKPLNLESEVSLTALTSQVNTSNTVQILLSKATMPKQKGVN
LVSTPEQHSQSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP
ERFLYCSQARSPARVNSKAPSVSTCVSPSPSPSPSPSPSPSPSPSPSPSP
PSKFAVAGDCPRLLSALAEQELGLQEAIVEKADKAWAASLPKNDGKPPF
SLENHACLTVAKTSWQSEPEPTPELDKGGSTKDFSPGLFSPVAFATDP
KTSISPSGTPLGTATPDPTAFDFCDPTASSVDGNAPWPEPNDGACPR
WGLHPEGLLQEQKASQKADHRAKACVGFQEDHAKGAALSDFKQEA
EGYKEEVLGLOCPKADQWLEESRCHSCSTDFGDLPLPPPKKEDLSAEELYS
LCELLSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP
VAGSKVQWFFSLSHMKPQSPMERAPGSGSTQSGSLAPKPNKPAVEPIAK
EPVPRGKLSRRVHRLPEAEDSPCRVAPLPPKLLPESCTPPQQAEGAGPGR
LSEGLPMKTRSLTALSEPDPGGLTPTTPPKLGGKORAAFKSKRVKPSRA
PPPPPLPAPACMLQSTQSLPQAVRTKLPVPPKGLKLEAIVQKITSPLKLLAC
RVAGPPTPRGALPERRPGSPAGAEGLGGQMLPAAGADPLCNPAASRLKG
KLLSKLSAADCPKAEAFWSPPTPLSGTARAPKRSKRTGTGLGSPGLEKRP
PGQFLLAHDRASSTGGEDNSGCKKPEELGPASQPPGRCPCOPCQRAQK
PQASYSYSKRKRSLRGRKTHASPGCKRAIRRRQQQVLLDPAEPRLKVISSC
KLRADSRTPAFSPVRYVEKDAITICTIVNSQDEPKPHWPPSSAASSTSSSL
BPAGASITTPPGSLVQSLTSMHLPVVSQKALSTSLVCLQOONANFKDLG
DLCGVPYPCILPKPKKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEK
ITTFALGRSLRDPADPAQKGLRTSARGLSRLQSCYCDGQDGGEEVAADKS
KHECSKEAETEGGDTQEHNVHACAVTWSGVILVAGKLFGLQKAMKVAVDMVRG"
7202..7207
polyA_signal
7208..7213
polyA_site
7222
BASE COUNT 1612 a 2335 c 2027 g 1248 t
ORIGIN
Query Match 51.6%; Score 170.2; DB 10; Length 7222;
Best Local Similarity 75.8%; Pred. No. 4,2e-24;
Matches 250; Conservative 0; Mismatches 49; Indels 31; Gaps 2;
QY 1 tcttccacactctccactactctctctctctccagagtggtggcaggggccna 60
|||||
Db 930 TCTTCCCTACCTCTCCACTATGACCCCAACAGT-GCAGGGTGTGGCAGGGGCCA 988
|||||
QY 61 ctctataagagttgcacagcagcagctgccagcgcctatgacagcgccgtgactgcag 120
|||||
Db 989 CTCCTACAGAGCTGCACAGCACCATCTGCCAGCCTCATGATAGGCCGATGATGCCAA 1048
|||||
QY 121 ctccagctgccccggcgaggggtccagaattcttccatccctacagctcgggcgct 180
|||||
Db 1049 TCGCAACCTGGCTCCAGGCAACGGGTCCAGAACTCTTCACGCTTACAGCGCTGGCGGCT 1108
|||||
QY 181 cagctatgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 240
|||||
Db 1109 TGGCTAGC-----AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1138
|||||
QY 241 gagcggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 300
|||||
Db 1139 AGCGCGTACACACACAGGAAACACTCCACTACAGAACCTTCGCCAAGTACCAACTA 1198
|||||
QY 301 cggggaagaagcaggggtactgcagcc 330
|||||
Db 1199 TGGACAGCAAGCGAGGGGTACTGTCCACC 1228
|||||
RESULT 12
AC073946/c
LOCUS AC073946 197232 bp DNA HTG 16-JAN-2001
```

DEFINITION Mus musculus chromosome 10 clone RP23-218e4, WORKING DRAFT  
SEQUENCE, 7 unordered pieces.

ACCESSION AC073946  
VERSION AC073946.14 GI:12232501  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 197232)  
AUTHORS Fu, Y., Wang, Q., Pan, H. and Roe, B. A.  
TITLE Mus musculus Chromosome 10 BAC Clone rp23-218e4  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 197232)  
AUTHORS Fu, Y., Wang, Q., Pan, H. and Roe, B. A.  
TITLE Direct Submission  
JOURNAL Submitted (07-JUL-2000) Department of Chemistry And Biochemistry,  
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA

COMMENT  
\* ON Jan 16, 2001 this sequence version replaced gi:11641343.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 7 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 2831: contig of 2831 bp in length  
\* 2931: gap of unknown length  
\* 2932 15879: contig of 12948 bp in length  
\* 15880 15979: gap of unknown length  
\* 15980 33201: contig of 17222 bp in length  
\* 33202 33301: gap of unknown length  
\* 33302 51575: contig of 18274 bp in length  
\* 51576 51675: gap of unknown length  
\* 51676 81126: contig of 29451 bp in length  
\* 81127 81226: gap of unknown length  
\* 81227 113588: contig of 32362 bp in length  
\* 113589 113688: gap of unknown length  
\* 113689 197232: contig of 83544 bp in length.

FEATURES  
Location/Qualifiers  
1..197232  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="10"  
/clone="RP23-218e4"  
BASE COUNT 49202 a 48272 c 48817 g 50340 t 601 others  
ORIGIN

Query Match 20.8%; Score 68.8; DB 2; Length 197232;  
Best Local Similarity 56.7%; Pred. No. 0.00016;  
Matches 127; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 92 cagccccatgacagcgccgtgactgcccagctccagctgccccggcgagcggtccag 151  
|||||  
Db 81609 CAGCACCAC 81550  
|||||  
QY 152 aattctcactcactcactcactcactcactcactcactcactcactcactcactc 211  
|||||  
Db 81549 CAC 81490  
|||||  
QY 212 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 271  
|||||  
Db 81489 CACCAGCACCAC 81430  
|||||  
QY 272 taccaaaactgcagagctatcagcactacgagcagcagcagcagcagcagcagc 315  
|||||  
Db 81429 CAGCACCAC 81386  
|||||

LOCUS	1968 bp	05-DEC-1996
MMT46463	1968 bp	05-DEC-1996
DEFINITION	Mus musculus glutamine repeat protein-1 mRNA, complete cds.	

[illegible]

[illegible]

Search completed; December 4, 2001, 16:09:27  
Job time: 10610 sec







FTT

```

T      mRNA          /rpt_type= "CAG"
T      T            2473..3093
T      T            /*tag= b
T      T            2473..3023
T      T            /*tag= c
T      T            3024..3093
T      T            /*tag= d
T      T            /product= "ppp2R2B"
T      T            /note= "protein_id= AAF74024.1"
X
X
X      WO200078943-A2.
X
X      28-DEC-2000.
X
X      16-JUN-2000; 2000WO-US40213.
X
X      18-JUN-1999; 99US-0140176.
X      (UWJO ) UNIV JOHNS HOPKINS.
X      Margolis RL, Ross CA, Holmes SE;
X      WPI; 2001-102718/11.
X      P-PSDB; AAB48340.
X
X      New polynucleotide for use in diagnosing spinocerebellar ataxia 12,
X      comprises a microsatellite marker having a variable number of CAG
X      trinucleotide repeats located on human chromosome 5q31-33 -
X
X      Disclosure; Page 21-23; 23pp; English.
X
X      The invention relates to a microsatellite marker located on human
X      chromosome 5q31-33 between markers D5S402 and WI-6763, comprising a
X      variable number of CAG trinucleotide repeats. The marker is useful for
X      diagnosing spinocerebellar ataxia 12 (SCA12) by determining the number of
X      CAG trinucleotide repeats in a 5'-untranslated region of an allele of
X      PP2RBbeta, a brain specific regulatory subunit of protein phosphatase
X      PP2A, where if a number greater than 40 is determined, an expansion which
X      is associated with SCA12 is indicated. The presence of expanded
X      trinucleotide repeats on chromosome 5q31-33 is useful for predicting or
X      diagnosing SCA12. Also provided are primers useful for amplifying the
X      microsatellite marker associated with SCA12 and for diagnosing SCA12 and
X      a probe useful for detecting expansions in a microsatellite marker
X      located on human chromosome 5q31-33. The present sequence represents
X      a PP2R2B DNA sequence.
X
X      Sequence 5120 BP; 1336 A; 1267 C; 1268 G; 1249 T; 0 other;
X

```

Query Match	20.4%	Score 67.2	DB 22	Length 5120
Best Local Similarity	56.9%	Prod. No. 1e-05		
Matches 123	Conservative	0	Mismatches 93	Indels 0
Gaps				
QY 114	cttcgagcttcagcctcgcccggggagcagcggtccagaattcttcgctaccagtcgg	173		
Db				
QY 2068	cagcgctccagcctctctgcagcagcagcagcagcagcagcagcagcagc	2127		
Db				
QY 174	gcgcgctcagctatgaccagcagcagcagcagcagcagcagcagcagcagc	233		
Db				
QY 2128	agcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	2187		
Db				
QY 234	cccttcagcgcgcgcacctgcgcagaaaccttcattaccaaaacctcgccagtc	293		
Db				
QY 2188	agcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	2247		
Db				
QY 294	agcactacgggcagcagcagccagggctactgcagc	329		
Db				
QY 2248	agcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	2283		
Db				

RESULT 6  
AAT33007  
ID AAT33007 standard; DNA; 10266 BP.

XX	AC	AAAT33007;	
XX	DT	23-OCT-1996 (first entry)	
XX	DE	Mouse SRY-related gene.	
XX	KW	Mouse; SRY; primer; PCR; polymerase chain reaction; amplification; probe;	
XX	KW	HMG box; human; bovine; sex; animal; birth; ds.	
XX	OS	Mus musculus.	
XX	EH	Key	Location/Qualifiers
XX	FT	CDS	7148..8335
XX	FT	/*tag= a	
XX	FT	/product= SRY-related protein	
XX	PN	JP08154685-A.	
XX	PN	18-JUN-1996.	
XX	PD		
XX	PE	30-NOV-1994; 94JP-0319525.	
XX	PR	30-NOV-1994; 94JP-0319525.	
XX	PA	(KACH-) KACHIKU JUSEIRAN ISHOKU GIKUTSU KENKYUKU.	
XX	WPI	WPI; 1996-336575/34.	
XX	DR	P-PSDB; AAN03474.	
XX	PT	Bovine and mouse Sry-related DNA - useful for detecting e.g. the sex	
XX	RT	of unborn animals	
XX	PS	Claim 1; Page 10-14; 21pp; Japanese.	
XX	CC	This is the nucleotide sequence of a gene encoding a mouse SRY-related	
XX	CC	protein. The gene was isolated from a mouse genomic library using a	
XX	CC	cDNA fragment amplified by primers AAT33009-10 as a probe. The screen	
XX	CC	isolated 4 EcoRI fragments of 2.3, 2.8, 3.5 and 1.5 kb covering the	
XX	CC	gene. Sequence analysis revealed a 240 bp HMG box sequence between bases	
XX	CC	7154-7393. Similarity with the human SRY HMG box sequence resulted in	
XX	CC	primers being generated to amplify the human SRY HMG box sequence for	
XX	CC	use as a probe to isolate the bovine SRY-related gene (AAT33008). The	
XX	CC	mouse and bovine genes are useful for determining the sex of an animal	
XX	CC	prior to birth.	
XX	SO	Sequence 102666 BP; 2958 A; 2202 C; 2228 G; 2878 T; 0 other;	

	Query_Match	19.8%	Score 65.2	DB 17	Length 10266	
	Best Local Similarity	54.6%	Pred. No. 3e-05			
	Matches: 130	Conservative	0	Mismatches 108	Indels	0
						Gaps 0
Qy	192	cagcccaatgacagcgcgtgactgcagcttcagccttgcggcccggggagcagcggttcacg	151			
Db	7838	cagcagcagcagcagcagcagcagcttcgatgccaccaccagcaagaagcagcaggttcac	7897			
Qy	152	aatcttcatgcctaccagtcgtggcgccctcagttatgccagcagcgagcagcagcagc	211			
Db	7898	gaccaccaccaccacacagcagcagcagtcgttcattgccaccacgagcagcagcagcag	7957			
Qy	212	cagcagcagcagcagcagcagcagccttcagcagcgcgcacacctgccagcaaaccttcac	271			
Db	7958	ttcatatgaccaccagcagcagcagcagtcatttcacgaccaccaccacgagcagcagcag	8017			
Qy	272	taccaaaacctgcgaagtcacgacactacggcgacagggccagggctactgcgcagc	329			
Db	8018	ttcatatgaccaccagcagcagcagcagtcgttcattgccaccaccacacgagcagcagc	8075			

RESULT 7  
AAQ20685  
ID AAQ20685 standard; DNA; 14704 BP.











GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2001, 14:13:57 ; Search time 102.2 Seconds  
(without alignments)

731.289 Million cell updates/sec

Title:

US-09-508-821b-2

Perfect score:

330  
1 tcttcccccacactctccacac.....ggccagggctactgcagcc 330

Sequence:

IDENTITY\_NUC

Scoring table:

Gapop 10.0 , Gapext 1.0

Searched:

351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters:

702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing:

Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Issued\_Patents\_NA.\*  
1: /cgn2.6/prodata/2/ina/5A.COMB.seq.\*  
2: /cgn2.6/prodata/2/ina/5B.COMB.seq.\*  
3: /cgn2.6/prodata/2/ina/5A.COMB.seq.\*  
4: /cgn2.6/prodata/2/ina/5B.COMB.seq.\*  
5: /cgn2.6/prodata/2/ina/PCRUS.COMB.seq.\*  
6: /cgn2.6/prodata/2/ina/backfillseq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	59.4	18.0	234	1	US-08-469-802B-3	Sequence 3, Appl
2	59.4	18.0	234	2	US-08-267-803B-3	Sequence 3, Appl
3	59.4	18.0	397	3	US-09-253-691-3	Sequence 3, Appl
4	59.4	18.0	477	4	US-09-135-994-1	Sequence 1, Appl
5	58.8	17.8	203	4	US-09-043-303-7	Sequence 7, Appl
6	58.2	17.6	3254	3	US-08-965-903B-1	Sequence 1, Appl
7	56.6	17.2	154	1	US-08-469-802B-6	Sequence 6, Appl
8	56.6	17.2	154	2	US-08-267-803B-6	Sequence 6, Appl
9	56.4	17.1	171	1	US-08-469-802B-5	Sequence 5, Appl
10	56.4	17.1	171	2	US-08-267-803B-5	Sequence 5, Appl
11	56.4	17.1	195	2	US-08-469-802B-2	Sequence 2, Appl
12	56.4	17.1	195	2	US-08-267-803B-2	Sequence 2, Appl
13	56.2	17.0	168	1	US-08-469-802B-4	Sequence 4, Appl
14	56.2	17.0	168	2	US-08-267-803B-4	Sequence 4, Appl
15	54.6	16.5	165	4	US-09-043-303-17	Sequence 17, Appl
16	54.4	16.5	2793	1	US-08-458-298-1	Sequence 1, Appl
17	54.4	16.5	2793	1	US-08-209-747-1	Sequence 1, Appl
18	53.4	16.2	3366	1	US-08-469-802B-1	Sequence 1, Appl
19	53.4	16.2	3366	2	US-08-267-803B-1	Sequence 1, Appl
20	53.4	16.2	10660	4	US-08-267-803B-8	Sequence 8, Appl
21	53.4	16.2	10660	2	US-09-041-886-16	Sequence 16, Appl
22	53.2	16.1	506	1	US-08-469-802B-7	Sequence 7, Appl
23	53.2	16.1	506	2	US-08-267-803B-7	Sequence 7, Appl
24	53	16.1	3376	3	US-08-320-559-29	Sequence 29, Appl
25	53	16.1	3376	3	US-08-545-860D-29	Sequence 29, Appl
26	53	16.1	3376	5	PCR-US94-0496-29	Sequence 29, Appl
27	51.8	15.7	10348	2	US-08-457-273B-41	Sequence 41, Appl

28	51.8	15.7	10348	3	US-08-556-419-13	Sequence 13, Appl
29	51.8	15.7	10348	4	US-09-041-886-14	Sequence 14, Appl
30	51.8	15.7	10366	1	US-08-246-982A-5	Sequence 5, Appl
31	51.8	15.7	10366	1	US-08-453-265-5	Sequence 5, Appl
32	49.6	15.0	543	6	5273901-6	Patent No. 5273901
33	49.6	15.0	3489	2	US-08-728-323A-1	Sequence 1, Appl
34	49.6	15.0	32207	2	US-08-770-379-20	Sequence 20, Appl
35	49.6	15.0	32207	4	US-08-757-669A-20	Sequence 20, Appl
36	49.4	15.0	2201	3	US-09-330-970-2	Sequence 2, Appl
37	49.4	15.0	3336	3	US-09-330-970-4	Sequence 4, Appl
38	49.2	14.9	3715	4	US-09-041-886-10	Sequence 10, Appl
39	49	14.8	4835	1	US-08-386-495-9	Sequence 9, Appl
40	49	14.8	4835	5	PCR-US96-02331-9	Sequence 9, Appl
41	48.2	14.6	1776	2	US-08-531-927B-1	Sequence 1, Appl
42	48.2	14.6	1776	4	US-09-041-886-12	Sequence 12, Appl
43	48.2	14.6	2950	5	PCR-US93-08386-7	Sequence 7, Appl
44	48.2	14.6	2968	5	PCR-US93-08386-1	Sequence 1, Appl
45	48.2	14.6	3771	1	US-08-185-432-3	Sequence 3, Appl

#### ALIGNMENTS

RESULT 1  
US-08-469-802B-3  
Sequence 3, Application US/08469802B  
Patent No. 5741645  
GENERAL INFORMATION:  
APPLICANT: Orr, Harry T.  
APPLICANT: Ratum, Laura P.W.  
APPLICANT: Chung, Ming-Yi  
APPLICANT: Zoghbi, Huda Y.  
TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia  
Patent No. 5741645  
TITLE OF INVENTION: Type 1 and Method for Diagnosis  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:  
ADDRESS: Mueller, Raasch, Gebhardt & Schnappach, P.A.  
STREET: 119 No. 5741645th Fourth Street, Suite 203  
CITY: Minneapolis  
STATE: MN  
COUNTRY: USA  
ZIP: 55401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: US/08/469,802B  
APPLICATION NUMBER: US/08/469,802B  
CLASSIFICATION: 435  
FILING DATE: 06-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Mueller, Ann M.  
REGISTRATION NUMBER: 33,977  
REFERENCE/DOCKET NUMBER: 110,00030101  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-305-1217  
TELEFAX: 612-305-1225  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 234 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-08-469-802B-3  
Query Match 18.0%; Score 59.4; DB 1; Length 234;  
Best Local Similarity 56.3%; Pred. No. 1.7e-05;  
Matches 11; Conservative 0; Mismatches 86; Indels 0; Gaps 0;



	Query Match	18.0%;	Score 59.4;	DB 2;	Length 234;	
	Best Local Similarity	56.3%;	Pred. No. 1.7e-05;			
	Matches 111;	Conservative	0;	Mismatches 86;	Indels	0; Gaps 0;
QY	134	ccggggcagcggttcagaattctatgcctaccagtccggtcgccgtcagcatgacacg	193			
Dd	25	CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	84			
QY	194	cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcaact	253			

```

RESULT      4
US-09-135-994-1
: Sequence 1, Application US/09135994A
: Patent No. 6380938
: GENERAL INFORMATION:
: APPLICANT: Ranum et al.
: TITLE OF INVENTION: SCAT GENE AND METHODS OF USE
: FILE REFERENCE: University of Minnesota
: CURRENT APPLICATION NUMBER: US/09/135,994A
: CURRENT FILING DATE: 1998-08-18
: EARLIER APPLICATION NUMBER: 60/056,170
: EARLIER FILING DATE: 1997-08-19
: NUMBER OF SEQ ID NOS: 14
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 477
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-135-994-1

```

Query Match	18.0%;	Score 59.4;	DB 4;	Length 477;
Best Local Similarity	55.6%;	Pred. No. 2e-05;		
Matches 114;	Conservative	0;	Mismatches 91;	Indels 0;
				Gaps 0;

[illegible]

RESULT 5  
US-09-043-303-7  
; Sequence 7, Application US/09043303

```

1  APPLICANT: TSUTSI, Shoji
2  APPLICANT: SANBEI, Kazuhiro
3  TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Ataxia Type 2 and
4  TITLE OF INVENTION: Primers Therefor
5  FILE REFERENCE: 0760-0241P
6  CURRENT APPLICATION NUMBER: US/09/043,303
7  CURRENT FILING DATE: 1998-05-18
8  EARLIER APPLICATION NUMBER: PCT/JP96/01999
9  EARLIER FILING DATE: 1996-07-18
10 NUMBER OF SEQ ID NOS: 17
11 SOFTWARE: PatentIn Ver. 2.0
12 SEQ ID NO 7
13 LENGTH: 203
14 TYPE: DNA
15 ORGANISM: p-2093 plasmid
16 US-09-043-303-7

```

Query Match	17.8;	Score 58.8;	DB 4;	Length 203;
Best Local Similarity	56.1;	Pred. No. 2.3e-05;		
Matches 11;	Conservative	0;	Mismatches 87;	Indels 0;
				Gaps 0;

[illegible]

```

RESULT 6
US-08-965-903B-1
; Sequence 1, Application US/08965903B
; Patent No. 6060275
; GENERAL INFORMATION:
; APPLICANT: Hachohen, Nir

```

APPLICANT: KRASHOV, Mark A.  
 TITLE OF INVENTION: SPROUTY PROTEIN AND CODING  
 TITLE OF INVENTION: SEQUENCE  
 NUMBER OF SEQUENCES: 20  
 CORRESPONDENCE ADDRESS:  
 ADDRESSER: Dellinger & Associates  
 STREET: 350 Cambridge Ave., Suite 250  
 CITY: Palo Alto  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 94306  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FASTSEQ for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/965,903B  
 FILING DATE: 07-NOV-1997  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/030232  
 FILING DATE: 07-NOV-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Petithory, Joanne R  
 REGISTRATION NUMBER: 42,995  
 REFERENCE/DOCKET NUMBER: 8600-0177.30  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 650-324-0880  
 TELEFAX: 650-324-0960  
 TELEX:  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3254 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 ORIGINAL SOURCE:  
 INDIVIDUAL ISOLATE: Drosophila spy cDNA  
 FEATURE:  
 NAME/KEY: Coding Sequence  
 LOCATION: 1323...3095  
 OTHER INFORMATION:  
 US-08-965-903B-1

Query Match	17.6%;	Score 58.2;	DB 3;	Length 3254;
Best Local Similarity	59.3%;	Pred. No. 5.4e-05;		
Matches 99;	Conservative 0;	Mismatches 68;	Indels 0;	Gaps 0;

QY	83	1965	143	2025	203	2085
	ccgacgctccgcagccccaatgaaagccgcgactgtgcagccgcagccgctgtgcccgcggagag	ccggctggccacagaggaataatggcagacagacacacccacacactgtgtgtgtgtcccccag	cggtgtcagagatcttcacagctccacgcctgcgtgcgcgtcagctctatgacacagcagagag	cggatattgacacactctccacccctcacaacacacagcagcactctgtaaacgacacacacacag	cagc	cagc
	111	111	111	111	249	2131
	2024	2024	2022	2084		

US-08-469-802B-6  
Sequence 6, Application US/08469802B  
Patent No. 5711645

GENERAL INFORMATION:

APPLICANT: Orr, Harry T.  
APPLICANT: Rannum, Laura P.W.  
APPLICANT: Chung, Ming-Yi  
APPLICANT: Zoghbi, Huda Y.

RESULT 9  
 US-08-469-802B-5  
 ; Sequence 5, Application US/08469802B  
 ; Patent No. 5741645  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ozr, Harry T.  
 ; APPLICANT: Rannum, Laura P.W.  
 ; APPLICANT: Chung, Ming-yi  
 ; APPLICANT: Zoghbi, Huda Y.  
 ; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia  
 ; Patent No. 5741645  
 ; TITLE OF INVENTION: Type 1 and Method for Diagnosis  
 ; NUMBER OF SEQUENCES: 47  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Mueeting, Raasch, Gebhardt & Schwappach, P.A.  
 ; STREET: 119 No. 5741645th Fourth Street, Suite 203  
 ; CITY: Minneapolis  
 ; STATE: MN  
 ; COUNTRY: USA  
 ; ZIP: 55401  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/469,802B  
 ; FILING DATE: 06-JUN-1995  
 ; CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: Muehling, Ann M.

REGISTRATION NUMBER: 33,977

REFERENCE/DOCKET NUMBER: 110.00030101

TELECOMMUNICATION INFORMATION:

TELEPHONE: 612-305-1217

TELEFAX: 612-305-1225

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 171 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

US-08-469-802B-5

## Query Match

17.1%; Score 56.4; DB 1; Length 171;

Best Local Similarity 59.3%; Pred. No. 7.7e-05;

Matches 96; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY

134 ccggggcagcggggtccatcttcacgagtcggcgccctcagctatgacag 193

DB

1 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 60

QY

194 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 253

DB

61 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 120

QY

254 gccacggaacccctcattaccacaaacctcgcaagatcag 295

DB

121 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 162

## RESULT 10

US-08-267-803B-5

Sequence 5, Application US/08267803B

Patent No. 5834183

GENERAL INFORMATION:

APPLICANT: Ott, Harry T.

APPLICANT: Rannum, Laura P. W.

APPLICANT: Chung, Ming-Yi

APPLICANT: Zoghbi, Huda Y.

TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia

Patent No. 5834183

TITLE OF INVENTION: Type 1 and Method for Diagnosis

NUMBER OF SEQUENCES: 85

CORRESPONDENCE ADDRESS:

ADDRESSER: Muehling, Raasch, Gebhardt &amp; Schwappach, P.A.

STREET: P.O. Box 581415

CITY: Minneapolis

STATE: MN

COUNTRY: USA

ZIP: 55458-1415

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/267,803B

FILING DATE: 28-JUN-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Muehling, Ann M.

REGISTRATION NUMBER: 33,977

REFERENCE/DOCKET NUMBER: 110.00030101

TELECOMMUNICATION INFORMATION:

TELEPHONE: 612-305-1217

TELEFAX: 612-305-1225

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 171 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

US-08-267-803B-5

Query Match

17.1%; Score 56.4; DB 2; Length 171;

Best Local Similarity 59.3%; Pred. No. 7.7e-05;

Matches 96; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY

134 ccggggcagcggggtccatcttcacgagtcggcgccctcagctatgacag 193

DB

1 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 60

QY

194 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 253

DB

61 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 120

QY

254 gccacggaacccctcattaccacaaacctcgcaagatcag 295

DB

121 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 162

RESULT 11

US-08-469-802B-2

Sequence 2, Application US/08469802B

Patent No. 5741645

GENERAL INFORMATION:

APPLICANT: Ott, Harry T.

APPLICANT: Rannum, Laura P. W.

APPLICANT: Chung, Ming-Yi

APPLICANT: Zoghbi, Huda Y.

TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia

Patent No. 5741645

TITLE OF INVENTION: Type 1 and Method for Diagnosis

NUMBER OF SEQUENCES: 47

CORRESPONDENCE ADDRESS:

ADDRESSER: Muehling, Raasch, Gebhardt &amp; Schwappach, P.A.

STREET: 119 No. 5741645th Fourth Street, Suite 203

CITY: Minneapolis

STATE: MN

COUNTRY: USA

ZIP: 55401

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/469,802B

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Muehling, Ann M.

REGISTRATION NUMBER: 33,977

REFERENCE/DOCKET NUMBER: 110.00030101

TELECOMMUNICATION INFORMATION:

TELEPHONE: 612-305-1217

TELEFAX: 612-305-1225

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 195 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

US-08-469-802B-2

## Query Match

17.1%; Score 56.4; DB 1; Length 195;

Best Local Similarity 58.2%; Pred. No. 7.9e-05;

Matches 99; Conservative 0; Mismatches 71; Indels 0; Gaps 0;



Search completed: December 4, 2001, 16:10:39  
Job time: 7002 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2001, 12:59:46 ; Search time 2862.31 seconds

(without alignments)  
1238.897 Million cell updates/sec

Title: US-09-508-821B-2

Perfect score: 330  
Sequence: 1 tctctccacactctctccac.....ggccagggcactgacgaccc 330

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 11351937 seqs, 537289281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estom:\*  
5: em\_estipl:\*  
6: em\_estiba:\*  
7: em\_estiro:\*  
8: em\_estrov:\*  
9: em\_hic:\*  
10: gp\_estl1:\*  
11: gp\_estl2:\*  
12: gp\_hic:\*  
13: gp\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_iny:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rtd:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Query	Score	Match	Length	DB	ID	Description
1	133.4	40.4	963	10	AI893920	mc85a07.y
2	128.2	38.8	214	11	BG990415	MR2-RT116
3	86.4	26.2	224	10	AM492162	UT-M-BH3-
4	86.4	26.2	452	10	BE651204	UT-M-BH3-
5	69.4	21.0	805	13	AZ837781	ZMO133011
6	68.8	20.8	805	13	CNS04RM2	AL104427 Tetradon
7	65.4	19.8	987	12	AK013560	Mus muscu
8	64	19.4	884	13	CNS02NX2	AL105967 Tetradon
9	63.4	19.2	831	10	BE705571	SC01_03b1
10	63.2	19.0	605	13	AZ640368	BE705571
11	62.8	19.0	246	13	AZ593051	IM0502E21
12	62.4	18.9	1021	13	CNS01G7A	IM0404E19
						AL143207 Anopheles

c 13	61.6	18.7	1019	13	CNS01OY	AL154154 Anopheles
c 14	61	18.5	452	10	AJ283665	AA3B-AAE-
c 15	61	18.5	648	10	AI257238	AI257238
c 16	60.4	18.3	526	13	AZ12601	LM0028104
c 17	60.2	18.2	572	13	CNS043SN	AL273200 Tetradon
c 18	60	18.2	628	11	BE630444	BE630444 602766990
c 19	60	18.2	2199	10	BE741896	BE741896 601595356
c 20	59.4	18.0	356	10	AV627762	AV627762
c 21	59.4	18.0	492	10	AA665310	BP230018B
c 22	59	17.9	1003	13	CNS04OZ	AL303092 Tetradon
c 23	58.8	17.8	555	13	AZ281461	AZ281461 RPI-23-1
c 24	58.8	17.8	562	13	AZ251135	AZ251135 RPI-23-5
c 25	58.8	17.8	806	13	CNS01JUL	AI109481 Drosophila
c 26	58.8	17.8	906	13	CNS021IL	AI199074 Tetradon
c 27	58.6	17.8	495	13	AO189056	AO189056 HS_3195_B
c 28	58.6	17.8	1007	13	CNS01M2O	AI151699 Anopheles
c 29	57.8	17.5	947	13	CNS011S7	AI146136 Anopheles
c 30	57.6	17.5	594	13	AZ870327	AZ870327 2M0182G13
c 31	57.6	17.5	980	13	CNS01BHD	AI154817 Anopheles
c 32	57.6	17.5	982	13	CNS06RC	AI411922 T3 end of
c 33	57.6	17.5	1096	11	BG173272	BG173272 60236762
c 34	57.6	17.5	1101	13	CNS016KS	AI106870 Drosophila
c 35	57.4	17.4	264	11	BE798211	BE798211 601585978
c 36	57.4	17.4	590	13	CNS03M2Z	AI250244 Tetradon
c 37	57.4	17.4	932	13	CNS01B9Z	AI141082 Anopheles
c 38	57.2	17.3	723	13	AZ627893	IM0463118
c 39	57	17.3	680	10	AI388163	AI388163 GI8971.5
c 40	56.8	17.2	351	13	FR0004775	Z88558 F.rubripes
c 41	56.8	17.2	526	13	CNS03CMO	AL237969 Tetradon
c 42	56.8	17.2	540	10	AA644269	AA644269 BP230015A
c 43	56.6	17.2	752	10	AA940845	AA940845 LD23113.5
c 44	56.6	17.2	927	10	CNS01JAS	AI146589 Anopheles
c 45	56.6	17.2	1002	13	CNS01N6I	AI151835 Anopheles

#### ALIGNMENTS

RESULT 1  
AI893920  
LOCUS  
DEFINITION  
AI893920  
Accession  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
COMMENT

AI893920 963 bp mRNA  
mc85a07.y1 Soares mouse embryo NBM13.5 14.5 Mus musculus cDNA  
clone IMAGE:355284 5' similar to gb:D29801 Mouse mRNA for unknown  
product, complete cds (MUSE);, mRNA sequence.  
AI893920  
GI:5599822  
EST.  
house mouse.  
Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 963)  
Marras, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Person,  
B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter,  
E., Kohn, S., Shinn, T., Jackson, T., Cardenas, M., McCann, R.,  
Waterston, R. and Wilson, R.  
The WashU-NCI Mouse EST Project 1999  
Unpublished (1999)  
Other ESTs: mc85a07.x1  
Contact: Marras M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63106, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LIND; contact the  
IMAGE Consortium (info@image.lind.gov) for further information.  
This read is a RESEQUENCE of a previously sequenced mouse clone  
This read has been verified (found to hit its original self in the  
correct orientation)  
MGI:227084  
Seq primer: -40RP from gibco







[illegible]

[illegible]



ORGANISM Mus musculus  
MUSCULUS  
Mammalia: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:  
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.  
1 (bases 1 to 246)  
Dunn, D., Aoyagi, A., Barber, M., Beecorn, T., Duval, B., Hamil, C.,  
Islam, H., Longacre, S., Maimoud, M., Meenen, E., Petersen, T., Rilly  
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.  
and Wright, D., Weiss, R.  
Mouse whole genome scaffolding with paired end reads from 10xb  
plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: dunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0404 row: E column: 19  
Seq primer: GGTGTAAACGACGCCCACT  
Class: Plasmid ends  
High quality sequence stop: 246.  
Location/Qualifiers  
1..246

location/Qualifiers  
1. 246  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="WUCM1M0404E19"  
/clone\_11b="Mouse 10kB plasmid WUCM1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-gold, T1-resistant, F-"  
/note="Vector: pMDA2n; Purified genomic DNA from M.  
laboratory mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptor DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pMD2 (g114732111g1AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. coli XL10-gold (Stratagene) cells  
and selected for ampicillin resistance."

[illegible]

[illegible]

RESULT 12
CNS01GJA/c
LOCUS
DEFINITION
from strain PEST of Anopheles gambiae (African malaria mosquito), genomic survey sequence.
ACCESSION
AL143207.1 GI:7001325
VERSION
KEYWORDS
SOURCE
African malaria mosquito.
ORGANISM
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anophelinae.
REFERENCE
1 (bases 1 to 1021)
Genoscope.
AUTHORS
Direct Submission
TITLE
Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segreff@genoscope.cns.fr)
JOURNAL
- Web : www.genoscope.cns.fr)
REFERENCE
2 (bases 1 to 1021)
AUTHORS
Roth,C.W., Brey,P.T., Ke.Z., Collins,F.H. and Weissenbach,J.
TITLE
Direct Submission
JOURNAL
Submitted (16-FEB-2000) BMJ, Institut Pasteur, 25, rue du Dr. Roux, Paris 75015, France
COMMENT
This clone is from an A. gambiae BAC library provided by F.H. Collins and sequenced by Genoscope in collaboration with the Laboratory of Biochem. and Biol. Molec. of Insects, Institut Pasteur.
FEATURES
Location/Qualifiers
1..1021
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="06C01"
/clone_lib="Notredame1"
/note="end : T7"
DATABASE COUNT 120 a 289 c 292 g 309 t 11 others
ORIGIN
Query Match 18.9%; Score 62.4; DB 13; Length 1021;
Best Local Similarity 54.3%; Pred. No. 0.007;
Matches 126; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
QY 85 gactgccagcccatgaacgcccgtgactgcagctccagcctgccccggcgacg 144
DB 955 GCAGCGCCAGGGGCCACAGCTCTCGMAACCCKGCCTGCCACAGGACACGACGCA 896
QY 145 ggtcagaattcttcatgctaccagtggcgccgtccatgctatgaccagcagcagca 204
DB 895 GGAGCGCGGCGAGCAGCACACATCAGCAGCGGGGAACAGCAGCAGCAGCAGCA 836
QY 205 gcagcagcagcagcagcagcagcagcagcccttcagagccgcgcacctgccccagaaac 264
DB 835 GCAGCAGCAGCAGCAGCAGCAGCAACAACACAGCAGCAGCGGAAACCCAGCAGCGTGAATG 776
QY 265 cctccattacaaaacctgcgaagtatcagcactacaggcagcagcagccag 316
DB 775 GCAGCAGCAGCAGCAGCAGCAGCATCAACAGCAGCGAGCAACAGCAGCAGCGAG 724
RESULT 13
CNS01OY/c
LOCUS
DEFINITION
Anopheles gambiae GSS SP6 end of clone 27H07 of Notredame1 library from strain PEST of Anopheles gambiae (African malaria mosquito), genomic survey sequence.
ACCESSION
AL154154
VERSION
ALI54154.1 GI:7015073
KEYWORDS
GSS.
SOURCE
African malaria mosquito.
ORGANISM
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anophelinae.
REFERENCE
1 (bases 1 to 1021)
Genoscope.
AUTHORS
Direct Submission
TITLE
Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segreff@genoscope.cns.fr)
JOURNAL
- Web : www.genoscope.cns.fr)
REFERENCE
2 (bases 1 to 1021)
AUTHORS
Roth,C.W., Brey,P.T., Ke.Z., Collins,F.H. and Weissenbach,J.
TITLE
Direct Submission
JOURNAL
Submitted (16-FEB-2000) BMJ, Institut Pasteur, 25, rue du Dr. Roux, Paris 75015, France
COMMENT
This clone is from an A. gambiae BAC library provided by F.H. Collins and sequenced by Genoscope in collaboration with the Laboratory of Biochem. and Biol. Molec. of Insects, Institut Pasteur.
FEATURES
Location/Qualifiers
1..1021
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="06C01"
/clone_lib="Notredame1"
/note="end : T7"
DATABASE COUNT 120 a 289 c 292 g 309 t 11 others
ORIGIN
Query Match 18.9%; Score 62.4; DB 13; Length 1021;
Best Local Similarity 54.3%; Pred. No. 0.007;
Matches 126; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
QY 85 gactgccagcccatgaacgcccgtgactgcagctccagcctgccccggcgacg 144
DB 955 GCAGCGCCAGGGGCCACAGCTCTCGMAACCCKGCCTGCCACAGGACACGACGCA 896
QY 145 ggtcagaattcttcatgctaccagtggcgccgtccatgctatgaccagcagcagca 204
DB 895 GGAGCGCGGCGAGCAGCACACATCAGCAGCGGGGAACAGCAGCAGCAGCAGCA 836
QY 205 gcagcagcagcagcagcagcagcagcagcccttcagagccgcgcacctgccccagaaac 264
DB 835 GCAGCAGCAGCAGCAGCAGCAGCAACAACACAGCAGCAGCGGAAACCCAGCAGCGTGAATG 776
QY 265 cctccattacaaaacctgcgaagtatcagcactacaggcagcagcagccag 316
DB 775 GCAGCAGCAGCAGCAGCAGCAGCATCAACAGCAGCGAGCAACAGCAGCAGCGAG 724
RESULT 13
CNS01OY/c
LOCUS
DEFINITION
Anopheles gambiae GSS SP6 end of clone 27H07 of Notredame1 library from strain PEST of Anopheles gambiae (African malaria mosquito), genomic survey sequence.
ACCESSION
AL154154
VERSION
ALI54154.1 GI:7015073
KEYWORDS
GSS.
SOURCE
African malaria mosquito.
ORGANISM
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anophelinae.
REFERENCE
1 (bases 1 to 1021)
Genoscope.
AUTHORS
Direct Submission
TITLE
Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segreff@genoscope.cns.fr)
JOURNAL
- Web : www.genoscope.cns.fr)
REFERENCE
2 (bases 1 to 1021)
AUTHORS
Roth,C.W., Brey,P.T., Ke.Z., Collins,F.H. and Weissenbach,J.
TITLE
Direct Submission
JOURNAL
Submitted (16-FEB-2000) BMJ, Institut Pasteur, 25, rue du Dr. Roux, Paris 75015, France
COMMENT
This clone is from an A. gambiae BAC library provided by F.H. Collins and sequenced by Genoscope in collaboration with the Laboratory of Biochem. and Biol. Molec. of Insects, Institut Pasteur.
FEATURES
Location/Qualifiers
1..1021
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/

**COMMENT**

Contact: Dimopoulos G  
Fotis C. Kafatos Laboratory  
European Molecular Biology Laboratory  
Meyerhofstrasse 1, 69117 Heidelberg, Germany  
f.kafatos@embl.org

**FEATURES**  
source

/organism="Anopheles gambiae"  
 /strain="4A r/r"  
 /db\_xref="taxon:7165"  
 /clone="4A35-A46-B-12"  
 /clone\_1b="Anopheles gambiae immune competent 4A35"  
 /cell\_line="immune competent 4A35"  
 /lab\_host="E. coli DH10B"  
 /note="Vector: pUT73D-pac (Pharmacia) with a modified  
 polylinker; Site.1: EcoRI; Site.2: NotI; sequenced from  
 the forward priming site that reads from the 3' end of  
 cDNA. The 4A35 is a directionally cloned and normalized  
 cDNA library that was constructed from the 4A35 cell line  
 o130-7 primed cDNA according to: Bonaldo, Lennon & Soares  
 (1996): Normalization and Subtraction: Two Approaches To  
 Facilitate Gene Discovery, Genome Research 6, 791-806."

Query Match	18.5%	Score 61;	DB 10;	Length 452;
Best Local Similarity	51.1%	Pred. No. 0.013;		
Matches 142; Conservative		0;	Mismatches 136;	Indels 0;
			Gaps	0.

Email: [http://www.fruitfly.org/EST\\_esterfruitfly.berkeley.edu](http://www.fruitfly.org/EST_esterfruitfly.berkeley.edu)  
 Plate: 55 row: D column: 10  
 High quality sequence stop: 432.  
 Location/Qualifiers  
     1..648  
         source  
             "Genbank:trioscobila melanogaster"

source

```

/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="LP05546"
/clone_1ib="LP Drosophila melanogaster larval-early pupal
por2"
/sex="male and female"
/dev_stage="larvae-pupae"
/lab_host="DHS-alpha"
/notes="Organ: whole body; Vector: por2; Site.1: EcoRI;
Site.2: XhoI; Sized fractionated cDNAs were directly
ligated into por2. Plasmid cDNA library."

```

Query Match	18.5%	Score 61;	DB 10;	Length 648;
Best Local Similarity	54.1%;	Pred. No. 0.013;		
Matches 124;	Conservative	0;	Mismatches 105;	Indels 0;
			Gaps	0

Search completed: December 4, 2001, 14:52:14  
Job time: 6748 sec

Result No.	Score	Query Match	Length	DB	ID	Description
1	6022	100.0	6022	20	AAK23983	Human hsc71 DNA, H
2	4627.6	76.8	5395	21	AAZ88975	Human HSBG1 cDNA.
3	1280	21.3	3422	21	AAC774855	Human ORFX ORF.10
4	714	11.9	2409	22	AAD05635	Human secreted pro
5	657.2	10.9	766	22	AAU05594	Human secreted pro
6	218.5	3.6	250	21	AAK57095	Human colon cancer
7	158	2.6	158	22	AAH33839	Human colon cancer
8	123.2	2.0	135	21	AAC21521	Human secreted pro
9	113.2	1.9	1004	20	AAK23982	Human GCT10004 cDNA
10	85.4	1.4	14704	13	AAQ20685	PKS 741 insert cont
11	82.2	1.4	10266	17	AAK33007	Mouse SRY-related

12	69.4	1.2	1655	22	AA122870
13	69.4	1.2	1635	22	AA168082
14	69.4	1.2	1655	22	AA108454
15	69.4	1.2	1973	22	AA133556
16	69.4	1.2	1973	22	AA134918
17	69.4	1.2	1973	22	AA103446
18	69.4	1.2	1973	22	AA121781
19	69	1.1	1075	22	AA147064
20	69	1.1	1075	22	AA107466
21	69	1.1	1403	22	AA118593
22	69	1.1	1403	22	AA133941
23	69	1.1	1403	22	AA102498
24	68.8	1.1	530	21	AA055931
25	68.4	1.1	5120	21	AA084577
26	66.8	1.1	3076	19	AAV43874
27	66.4	1.1	477	21	AA444307
28	64.6	1.1	397	20	AA489891
29	64.4	1.1	1337	20	AA015766
30	63.4	1.1	1848	22	AA050585
31	62.6	1.0	7680	22	AA059276
32	62.2	1.0	2744	16	AA098478
33	62.2	1.0	446	22	AA123276
34	61.8	1.0	446	22	AA185856
35	61.8	1.0	446	22	AA108921
36	61.4	1.0	234	16	AA084832
37	61.4	1.0	1159	21	AA059240
38	61.2	1.0	1437	21	AA059242
39	61.2	1.0	1472	21	AA059241
40	60	1.0	114955	20	AA059241
41	59.6	1.0	234	21	AA045758
42	58.8	1.0	203	19	AAV30271
43	58.8	1.0	203	19	AAV17226
44	58.2	1.0	3254	19	AAV34402
45	57.6	1.0	135	16	AA084833

## ALIGNMENTS

Probe #12713 used t  
Probe #16768 used t  
Probe #8445 used t  
Probe #3489 for ge  
Probe #3604 used t  
Probe #4437 used t  
Probe #11714 for g  
Probe #15750 used t  
Probe #4755 used t  
Probe #526 for ge  
Probe #627 used t  
Probe #489 used t  
Probe #4489 used t  
Enucalyptus girards  
DNA sequence of hu  
Receptor type tyro  
Human SCAT7 genom  
Spherocebellar at  
Human gene express  
DNA sequence of an  
Human polynucleoti  
MISPL-containing p  
Probe #13209 for g  
Probe #17282 used t  
Probe #8912 used t  
Spherocebellar at  
An EcoRI fragment  
Exons E, C, B and  
Exons D, C, B and  
Human adenosine A1  
Polymorphic region  
Glutamine rich reg  
SCA2 gene CAG repe  
Drosophila melanoza  
Spherocebellar at

RESULT	1
AXX23983	
ID	AXX23983 standard; DNA; 6022 BP.
XX	
XX	
XX	AXX23983;
XX	
XX	25-JUN-1999 (first entry)
XX	
XX	Human hGTL DNA.
XX	
XX	Polymorphic CAG repeat; hGTL; diagnosis; prognosis; schizophrenia;
XX	human; transcription factor; neuroleptic activity; affective disorder;
XX	manic depression; neurodevelopmental brain disease; detection;
XX	phenotypic variability; ss.
XX	
XX	Homo sapiens.
XX	
XX	W09915639-A1.
XX	
XX	01-APR-1999.
XX	
XX	13-SEP-1998; 98MO-CA00884.
XX	
XX	19-SEP-1997; 97CA-2216057.
XX	
XX	(UWMC-) UNIV MCGILL.
XX	
XX	Sankelilat C, Joobor R, Rouleau GA;
XX	
XX	WPI; 1999-254703/21.
XX	
XX	A human GRI gene containing a transcribed polymorphic CAG repeat,
XX	found in the diagnosis and treatment of schizophrenia



Claim 1; Fig 4A-E; 41pp; English.

This invention describes novel human GRI (hGRI) transcription factor gene with neuroleptic activity containing a transcribed polymorphic CAG repeat. Allelic variants of the hGRI gene CAG repeat are associated with schizophrenia, affective disorders (especially manic depression), neurodevelopmental brain diseases or with phenotypic variability with respect to long term response to neuroleptic medication. Short (171-177 bp) allelic variants of CAG repeats in the hGRI gene, are indicative of non-severe schizophrenia and neuroleptic response in patients. Probes and/or primers designed using the hGRI gene can be used to identify genes interacting with a biochemical pathway affected by the hGRI gene. The identified gene role can then be evaluated in psychiatric patients. Therapeutic agents can be identified by administering the agent to a transgenic mammal (or schizophrenic patients) and evaluating the prevention and/or treatment of development of schizophrenia.

6022 BP: 1247 A: 2060 C: 1799 G: 916 T: 0 other;

Query Match	100.0%	Score 6022;	DB 20;	Length 6022;
-------------	--------	-------------	--------	--------------

Query Index: Best Local Similarity 100.0%; Pred. No. 0;

Matches 6022; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1	gga	ccag	cagg	ccca	aggga	gag	ggag	cgga	aaat	gtct	gt	ctaa	gt	cttt	ggc	60
1	gga	ccag	cagg	ccca	aggga	gag	ggag	cgga	aaat	gtct	gt	ctaa	gt	cttt	ggc	60
61	tgt	cag	aa	gggt	ggg	agt	gat	gggt	ggg	gaa	catt	ggg	gag	gag	cgga	120
61	tgt	cag	aa	gggt	ggg	agt	gat	gggt	ggg	gaa	catt	ggg	gag	gag	cgga	120
121	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	180
121	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	ggg	180
181	cag	tgg	cgc	tcct	ggc	aa	gact	cag	ctg	cag	gcg	atg	tgg	gag	cgga	240
181	cag	tgg	cgc	tcct	ggc	aa	gact	cag	ctg	cag	gcg	atg	tgg	gag	cgga	240
241	cac	tc	ccct	gac	aca	ga	agt	tgt	ca	at	tgc	aca	gct	ggt	ggg	300
241	cac	tc	ccct	gac	aca	ga	agt	tgt	ca	at	tgc	aca	gct	ggt	ggg	300
301	ggg	gac	t	aa	ga	ctc	gcg	ggg	gag	cag	gcg	ggg	tgg	ggg	ggg	360
301	ggg	gac	t	aa	ga	ctc	gcg	ggg	gag	cag	gcg	ggg	tgg	ggg	ggg	360
361	ctt	ct	gagg	ca	aa	gga	agt	ggc	ctg	ct	ga	at	gct	ct	gc	420
361	ctt	ct	gagg	ca	aa	gga	agt	ggc	ctg	ct	ga	at	gct	ct	gc	420
421	aat	tc	ccct	ccct	tc	ct	tc	tc	tc	tc	tc	tc	tc	tc	tc	480
421	aat	tc	ccct	ccct	tc	ct	tc	tc	tc	tc	tc	tc	tc	tc	tc	480
481	gcc	gag	tac	gac	gtt	ttc	gaa	aa	gg	tgt	ggt	tt	ccat	ggc	aa	540
481	gcc	gag	tac	gac	gtt	ttc	gaa	aa	gg	tgt	ggt	tt	ccat	ggc	aa	540
541	cag	cag	ac	ctc	gcg	aa	aa	ca	tac	gc	ct	tag	aa	tta	cag	600
541	cag	cag	ac	ctc	gcg	aa	aa	ca	tac	gc	ct	tag	aa	tta	cag	600
601	cta	gct	g	ca	ccg	ca	gcg	gt	g	ct	cgc	aa	gac	t	at	660
601	cta	gct	g	ca	ccg	ca	gcg	gt	g	ct	cgc	aa	gac	t	at	660
661	ac	t	at	ca	ggt	gg	cg	ct	gg	cac	gc	ct	ct	g	ca	720
661	ac	t	at	ca	ggt	gg	cg	ct	gg	cac	gc	ct	ct	g	ca	720

1801 ctgacatctacagatgagaaacatctccaaacacggttcaagagatggtctctccaaagct 1860  
1861 gctgtgacgagaaagagatggttcaagaaactctgtctccaaagaccccaagagacataaa 1920  
1861 gctgtgacgagaaagagatggttcaagaaactctgtctccaaagaccccaagagacataaa 1920  
1921 agccagacatgacagcccaagagagagagctacatcaagccagacccgagagacacccgctg 1980  
1921 agccagacatgacagcccaagagagagagctacatcaagccagacccgagagacacccgctg 1980  
1981 tcaagagacgagacagacagcccaagacacagctccaaagacatgagagagagagagagac 2040  
1981 tcaagagacgagacagacagcccaagacacagctccaaagacatgagagagagagagagac 2040  
2041 ctgagagagaccccaagagagacacacatgagagagagacatctccaaagacccgctg 2100  
2041 ctgagagagaccccaagagagacacacatgagagagagacatctccaaagacccgctg 2100  
2101 agccctgagacagatgacaaagacacatgagagagagacacccgagacacccgctg 2160  
2101 agccctgagacagatgacaaagacacatgagagagagacacccgagacacccgctg 2160  
2161 gtagacatctctgagacagacatgacacacacacacacacacacacacacacacacacac 2220  
2161 gtagacatctctgagacagacatgacacacacacacacacacacacacacacacacacac 2220  
2221 agtctgacgac 2280  
2221 agtctgacgac 2280  
2281 ctgctgagacgac 2340  
2281 ctgctgagacgac 2340  
2341 atcgtgtgagagagac 2400  
2341 atcgtgtgagagagac 2400  
2401 aagac 2460  
2401 aagac 2460  
2461 tggac 2520  
2461 tggac 2520  
2521 aatgac 2580  
2521 aatgac 2580  
2581 gac 2640  
2581 gac 2640  
2641 gac 2700  
2641 gac 2700  
2701 gac 2760  
2701 gac 2760  
2761 tggagagagac 2820  
2761 tggagagagac 2820  
2821 ggcac 2880  
2821 ggcac 2880  
2881 cccac 2940  
2881 cccac 2940

2881 cccac 2940  
2881 cccac 2940  
2941 gtag 3000  
2941 gtag 3000  
3001 ctgag 3060  
3001 ctgag 3060  
3061 cccac 3120  
3061 cccac 3120  
3121 ctgag 3180  
3121 ctgag 3180  
3181 atctgac 3240  
3181 atctgac 3240  
3241 tggac 3300  
3241 tggac 3300  
3301 gtag 3360  
3301 gtag 3360  
3361 gtag 3420  
3361 gtag 3420  
3421 aagac 3480  
3421 aagac 3480  
3481 ttag 3540  
3481 ttag 3540  
3541 ctag 3600  
3541 ctag 3600  
3601 atgag 3660  
3601 atgag 3660  
3661 cgttct 3720  
3661 cgttct 3720  
3721 cctgac 3780  
3721 cctgac 3780  
3781 gtag 3840  
3781 gtag 3840  
3841 gtag 3900  
3841 gtag 3900  
3901 gtag 3960  
3901 gtag 3960  
3961 tccag 4020  
3961 tccag 4020

[illegible]



Y 2050 tccgaggacccatcgagcgagcttctctactgcaacacgagcccgctggcagccctgccc 2109  
b 1420 tccgaggacccatcgagcgagcttctctactgcaacacgagcccgctggcagccctgccc 1479  
Y 2110 agggctcaacagcaactcgagagcaacgcccagctcgctgctccactgtctctgacactc 2169  
b 1480 agggctcaacagcaactcgagagcaacgcccagctcgctgctccactgtctctgacactc 1539  
Y 2170 cctgacgacatgtccacacaaatctgagcactcctccagagcctccagagcctccagagcctccg 2229  
b 1540 cctgacgacatgtccacacaaatctgagcactcctccagagcctccagagcctccagagcctccg 1599  
Y 2230 ctcgacgactctccaaagtctcgagcgagcgagcgagctcgctcgctccactgtctctcagc 2289  
b 1600 ctcgacgactctccaaagtctcgagcgagcgagcgagctcgctcgctccactgtctctcagc 1659  
Y 2290 gccctggcagagagagcactggctccagagactcctgggctgaggaagcactcggctgag 2349  
b 1660 gccctggcagagagagcactggctccagagactcctgggctgaggaagcactcggctgag 1719  
Y 2350 aagggcgacaaagtgggctgaagcaccagcctgctcagagagcagcagcagcagcagcagc 2409  
b 1720 aagggcgacaaagtgggctgaagcaccagcctgctcagagagcagcagcagcagcagcagc 1779  
Y 2410 ttctcgtggagacacacagcgctgctgactctgtggcagagagtgctggcccccgg 2469  
b 1780 ttctcgtggagacacacagcgctgctgactctgtggcagagagtgctggcccccgg 1839  
Y 2470 cctggggagcgagcgccctccagacactctcagctggaacagggcgaggaagcagcagc 2529  
b 1840 cctggggagcgagcgccctccagacactctcagctggaacagggcgaggaagcagcagc 1899  
Y 2530 gacttcagccagcggtgtttgaagacccctcgtgctgctgctcagcagcagcagcagc 2589  
b 1900 gacttcagccagcggtgtttgaagacccctcgtgctgctgctcagcagcagcagcagc 1959  
Y 2590 aagacaaactgtctctctctctgttgacacagccactgtggctgctcagcagcagcagc 2649  
b 1960 aagacaaactgtctctctctctgttgacacagccactgtggctgctcagcagcagcagc 2019  
Y 2650 cccactacagcagctttgactgtttccggagacacacgctgcagcagcagcagcagc 2709  
b 2020 cccactacagcagctttgactgtttccggagacacacgctgcagcagcagcagcagc 2079  
Y 2710 gccaccccttgcctgcccagaggaacaaactggggatgttctccaggtggggattg 2769  
b 2080 gccaccccttgcctgcccagaggaacaaactggggatgttctccaggtggggattg 2139  
Y 2770 caccctggcagcttaccagggcctgagcagctggaagcctcagatgcatcagc 2829  
b 2140 caccctggcagcttaccagggcctgagcagctggaagcctcagatggaatcagc 2199  
Y 2830 aaaggggacacccatgagcttgcctgctgctgctgctgctgctgctgctgctgctgctg 2889  
b 2200 aaaggggacacccatgagcttgcctgctgctgctgctgctgctgctgctgctgctgctg 2259  
Y 2890 gagaagctgctcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 2949  
b 2260 gagaagctgctcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 2319  
Y 2950 gaggcagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 3009  
b 2320 gaggcagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 2379  
Y 3010 agccgagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 3069  
b 2380 agccgagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 2439  
Y 3070 aggaagggagcctggaagctgagagagagctcctcctcctgagctcctgagcagcagc 3129  
b 2440 aggaagggagcctggaagctgagagagagctcctcctcctgagctcctgagcagcagc 2499  
Y 3130 cccgagcagagcgctggcagtcagagaccccgctgcaacacagggcccccactcctctgccc 3189

OY		cccaagcaccacaaggaaagacctctccctctgactctcttctaatactctgtgccttcctc	5406
OY	5347	ccccagccccaagaaggaaagacctctccctctgactctcttctaatactctgtgccttcctc	5406
Db	4325	cccgaagccccaagaaggaaagacctctctctctctgctccctcttctaatactctgtgccttc	4384
OY	5407	tcccttgatcatcgacggcgagcgctccctcgcgcacacatccctggagcgctccatccctgcagcg	5466
Db	4585	tcccttgatcatcgacggcgagcgctccctcgcgcacacatccctggagcgctccatccctgcagcg	4644
OY	5467	cgggcctctctggccctctccctctcccaagaatgcacattggggcgctgtgtttccaagccctg	5526
Db	4645	cgggcctctctggccctctccctctcccaagaatgcacattggggcgctgtgtttccaagccctg	4704
OY	5527	agtaacctcttgacctgttgtctgtgctctctgcaccaaaaacccggccaatcttaagagaccttg	5586
Db	4705	agtaacctcttgacctgttgtctgtgctctctgcaccaaaaacccggccaatcttaagagaccttg	4764
OY	5587	gaacctctctgggccctactaacctccatgcacactctcccccgaagaagaacccaaatcccaag	5646
Db	4765	gaacctctctgggccctactaacctccatgcacactctcccccgaagaagaacccaaatcccaag	4824
OY	5647	gaaagaagctggcgccgaagaagaagcacctctggaagagcgctccctgcctgtttgagaagaactc	5706
Db	4825	gaaagaagctggcgccgaagaagaagcacctctggaagagcgctccctgcctgtttgagaagaactc	4884
OY	5707	aagagctcccgagctgtgacagtccgcgaactgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	5764
Db	4885	aagagctcccgagctgtgacagtccgcgaactgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	4944
OY	5765	gctgaacccggccaagcagcgcccaattgtagcaacatfgacccggcgccctgcgcgcgcgcgc	5824
Db	4945	gctgaacccggccaagcagcgcccaattgtagcaacatfgacccggcgccctgcgcgcgcgcgc	5004
OY	5825	cagagctcgcctactgctctgtgatccggagagatgtgtgtgtgtgtgtgtgtgtgtgtgtgt	5884
Db	5005	cagagctcgcctactgctctgtgatccggagagatgtgtgtgtgtgtgtgtgtgtgtgtgtgt	5064
OY	5885	aagagctgcgaacaacatgtagatcagcaaagagctcgcgcgcgcgcgcgcgcgcgcgcgcgc	5944
Db	5065	aagagctgcgaacaacatgtagatcagcaaagagctcgcgcgcgcgcgcgcgcgcgcgcgcgc	5069
OY	5945	gagcaactcgggtgatgaaagcgctctgtgcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	6004
Db	5070	--acactcgggtgatgaaagcgctctgtgcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	5127
OY	6005	aagctccttgtagctgcag	6022
Db	5128	aagctccttgtagctgcag	5145
 RESULT 3 AACC74855 ID AACC74855 standard; cDNA; 3422 BP.			
 AAC74855; DP -PDB-2001 (first entry)			
 Human ORFX ORF410 polynucleotide sequence SEQ ID NO:819.			
KW	Human; open reading frame; ORFX; detection: cytostatic; hepatotropic;		
KV	vulnerary; antipsoptic; antiliparxinsolan; nootropic; neuroprotective;		
KW	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;		
KW	immunostimulant; thrombolytic; coagulant; vasotropic; autidiabetic;		
KW	hypertensive; dermatological; immunosuppressive; antihematologic; antithyroid;		
KW	antiviral; antibacterial; antifungal; antihematuric; hypertension;		
KW	antihaemic; gene therapy; cancer; proliferative disorder; hyperextension;		
KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;		
KW	cardiovascular disease; diabetes mellitus; hypothyroidism; AIDS;		
KW	cholesterol ester storage; systemic lupus erythematosus; infection;		
KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;		
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;		
KW	bone damage; cartilage damage; antinflammatory disease; coagulation;		





	CC	gastrointestinal disorders, pregnancy-related disorders, endocrine disorders, and infectious. The proteins can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues, to identify their cognate ligands or binding partners, and in chemotaxis, and can be used as a food additive or preservative to modify storage properties.
	CC	Antibodies specific for a protein of the invention can be used in alleviating symptoms associated with the disorders mentioned above, and CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The present sequence represents a human secreted protein-encoding cDNA of the invention.
XX	Sequence	2409 BP; 490 A; 751 C; 711 G; 450 T; 7 other;
SQ	Query Match	11.9%; Score 714; DB 22; Length 2409;
	Best local similarity	99.5%; Pred. No. 3,9e+128;
	Matches 726:	Conservative 0; Mismatches 2; Indels 2; Gaps 1
Dy	5295	gcgaagacggtttacaacacatctgttaactccctcggagaagtgcaccaagcc 5354 
Db	12	gcgagacgcgttaccacacatacgactgttcacaccccgcggagatgcgccaaagc 71 
Dy	5355	ccaacaggagagcttctctctctgcctccctccctcaacctgcgtctcgttctcttgga 5414 
Db	72	cacacaggagagcttctctctctctctctctctctctctctccatccctcgtctctcttga 131 
OY	5415	tgcagdcgaggcctccctctgccccaatccctgagagctccatctctcacgcggcgccctc 5474 
Db	132	tgaagcggggcctccctctgccccaacctctggaggtcccatctctccagccggccctc 191 
OY	5475	cttgcgcctctctccacgatccattgggggccttgttttccaaagcccttagtacctc 5534 
Db	192	cttgcgcctctctccacgatccattgggggccttgtgtttccaaagcccttagtacctc 251 
OY	5535	ttcctctgttgcctgcctctgccccaaaacccggccaattcaaagaccttggggacctctg 5594 
Db	252	ttcctctgttgcctgcctctctgccccaaaacccggccaattcaaagaccttggggacctctg 311 
OY	5595	tgggacctctactaccctctaacactgctcccccacaaagaaagcacaactaaagaaaggtc 5654 
Db	312	tgggacctctactaccctctaacactgctcccccacaaagaaagcacaactaaagaaaggtc 371 
OY	5655	gcggccacagaagacacctgttagagagcctctgcctgcctgttagaggaacctacaaggtcc 5714 
Db	372	gcggccacagaagacacctgttagagagcctctgcctgcctgttagaggaacctacaaggtcc 431 
OY	5715	cgaatgtgcacgtctgcgcacactgcgcgggaagaccccccaagg--tgaagggccagctgaccc 5772 
Db	432	cgaatgtgcacgtctgcgcacactgcgcgggaagaccccccaagctcctgaagcccaagctgaccc 491 
OY	5773	ggccaaagcagggcccatctgcgcacagtgccccgggggacctctccgggggggtctgaaaggtctg 5832 
Db	492	ggccaaagcagggcccatctgcgcacagtgccccgggggacctctccgggggggtctgaaaggtctg 551 
OY	5833	ctactcgtctgtgatggccggagagatatggggcgagagagagagccccaagccgaagaggtctg 5892 
Db	552	ctactcgtctgtgatggccggagagatatggggcgagagagagagccccaagccgaagaggtctg 611 
OY	5893	caaaatctgactgacagagggggtctccggacagagcccggtgggggtggcccaaggaagactg 5952 
Db	612	caaaatctgactgacagagggggtctccggacagagcccggtgggggtggcccaaggaagactg 671 
OY	5953	ggtctgacagagcctctgtgcctgtgtggccggcgcgctctaacctgtgttggccggagaaagctct 6012 
Db	672	ggtctgacagagcctctgtgcctgtgtgggaacggcgcgctctaacctgtgttggccggagaaagctct 731 
OY	6013	tgggctgcag 6022 
Db	732	tgggctgcag 741 



5  
 005594  
 AAD05594 standard; cDNA; 766 BP.  
 AAD05594;  
 17-JUL-2001 (first entry)  
 Human secreted protein-encoding gene 16 cDNA clone HDPXY01, SEQ ID NO:26.26.  
 Human; secreted protein; proliferative disorder; cancer; tumour;  
 foetal abnormality; developmental abnormality; haematopoietic disorder;  
 immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 inflammation; allergy; neurological disorder; Alzheimer's disease;  
 Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
 skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
 cardiovascular disorder; angionenic disorder; kidney disorder;  
 gastrointestinal disorder; pregnancy-related disorder; gene therapy;  
 endocrine disorder; infection; wound healing; vulvuary;  
 cell culture; chemotaxis; food additive;  
 binding partner identification; ss

foetal abnormality; developmental anomalies; immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis; inflammation; allergy; neurological disorder; Alzheimer's disease; Parkinson's disease; cognitive disorder; schizophrenia; asthma; skin disorder; psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder; angiotensin disorder; kidney disorder; gastrointestinal disorder; pregnancy-related disorder; gene therapy; endocrine disorder; infection; wound healing; vulvure; cell culture; chemotaxis; food additive; banding partner identification; ss

Parkinson's disease; cognitive disorder; schizophrenia; asomnia;  
 skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
 cardiovascular disorder; angiodenic disorder; kidney disorder;  
 gastrointestinal disorder; pregnancy-related disorder; gene therapy;  
 endocrine disorder; infection; wound healing; vulnery;  
 cell culture; chemotaxis; food additive;  
 binding partner identification; ss

cardiovascular disorder; angiogenic disorder; kidney disorder; gastrointestinal disorder; pregnancy-related disorder; gene therapy; endocrine disorder; infection; wound healing; vulvuary; cell culture; chemotaxis; food additive; binding partner identification; ss

endocrine disorder; infection; wound healing; vulnery;  
cell culture; chemotaxis; food additive;  
binding partner identification; ss.

Homo sapiens.

[illegible]

```

CDS
23..319
/*tag= a

```

```

T      /product= "Human secreted protein
T      23..133
T      sig_peptide

```

```

T T
mat_peptide
/*tag= b
134..316
/*tag= c

```

```

/*tag= C
/product= "Human mature secreted protein"

```

WO200134627-A1.

X  
D  
V  
17-MAY-2001.

08-NOV-2000; 2000WO-US30628.

12-NOV-1999; 99US-0164744,  
30-JUN-2000; 2000US-0215140.

(HUMA-) HUMAN GENOME SCI INC.

Ruben SM, Komatsoulis GA, Baker KP, Young PE;

KX  
WPI; 2001-316491/33.

OR P-PSDB; AAE01785, encoding human secreted proteins, used in

New nucleic acid molecules encoding human sequences for preventing, treating or ameliorating a disorder, e.g. Alzheimer's and cancers -

XX  
XX  
pt parkinson's diseases and cancers  
XX Page 431. 55700. English

ps Claim I; Page 451; 30/07/97, England.

CC AAD03379-AAD03380 represent protein genes and AAE01770-AAE01849 represent the proteins they encode  
CC AAE01850-AAE01860 represent human secreted protein fragments or variants

The genes and their secreted proteins are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene

amount of the new protein in a sample or by determining the presence of

mutations in the new genes. Specific uses are described for each of the 28 genes, based on the tissues in which they are most highly expressed, and on the tissues in which they are induced or treatment of

CC and include developing products for the diagnosis or treatment of  
CC proliferative disorders, cancer, tumours, foetal and developmental  
CC diseases of the immune system.

CC abnormalities, haematopoietic disorders, diseases of the immune system.

CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation, rheumatism, rheumatoid arthritis, rheumatoid disease.

CC allergies, neurological disorders (e.g., Alzheimer disease),  
CC parkinson's disease), cognitive disorders, schizophrenia, asthma,  
CC parkinson's disease), cognitive disorders, schizophrenia, asthma,

CC skin disorders (e.g., psoriasis), sepsis, cancer, trauma  
CC cardiovascular disorders, angiogenic disorders, kidney disorders,

Db 738 ccaagccactaac 250

RESULT 7  
AAH33839/c

ID AAH33839 standard; cDNA; 158 BP.

AC AAH33839;

XX 03-SEP-2001 (first entry)

DE Human colon cancer antigen encoding cDNA SEQ ID NO:95.

XX Human: colon cancer; colon cancer antigen; diagnosis; detection;

KM colorectal carcinoma; SS.

XX Homo sapiens.

XX W0200122920-A2.

PN 05-APR-2001.

XX 28-SEP-2000; 2000NC-US26524.

XX 29-SEP-1999; 99US-0157137.

PR 03-NOV-1999; 99US-0163280.

XX (HMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Barash SC, Birse CE, Rosen CA.

PI WPI: 2001-235357/24.

DR P-PSDB; AAG74408.

PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
useful for preventing, diagnosis and/or treating colorectal cancers -

XX Claim 1; Page 2817; 9803pp: English.

XX AAH32943 to AAH37195 and AAG75514 to AAG77788 represent human colon  
cancer-associated nucleic acid molecules (N) and proteins (P), where  
the proteins are collectively known as colon cancer antigens. The colon  
cancer antigens have cytostatic activity and can be used in gene  
therapy and vaccine production. N and P may be used in the prevention,  
diagnosis and treatment of diseases associated with inappropriate P  
expression. For example, N and P may be used to treat disorders  
associated with decreased expression by rectifying mutations or deletions  
in a patient's genome that affect the activity of P by expressing  
inactive proteins or to supplement the patient's own production of P.  
Additionally, N may be used to produce the colon cancer-associated P,  
by inserting the nucleic acids into a host cell and culturing the cell  
to express the proteins. N and P can be used in the prevention, diagnosis  
and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
represent invention.

XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
missing at time of publication, meaning no sequences are present for  
SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 158 BP; 20 A; 45 C; 63 G; 30 T; 0 other;

XX

Query Match 2.6%; Score 158; DB 22; Length 158;  
Best Local Similarity 100.0%; Pred. No. 1.5e-21;  
Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps

XX 3652 atgtctatcgtctctctcctcaagcctgagccctcagcccgacgcccagcccgagcccgagcctg 3711

XX 158 AATGTGACTCTCTCTCTCAAGCCCTGATGAGCCCGCCGACCGCCGAGCCCGCAGGCTCG 99

XX 3712 acacgaccctcaccacccagagaaacggggggaagagagcgagccgcttcacgctc 3771

XX 98 ACCGACACCCCTGACCCCTCCAGAAACCTGGGGGGGCAAGCAAGGAGCGCGCTTCAAGTCG 39





```

Query Match      1.4%; Score 82.2; DB 17; Length 10266;
Best Local Similarity 46.1%; Pred. No. 1.2e-06;
Matches 276; Conservative 0; Mismatches 323; Indels 0; Gaps 0;

905 ccaagtatgatgaaacttgataaagacagcagctgcccccagcagcagcagatgcag 964
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7521 ccaaaacacgaagcgtttttattggcagcgttgatccccactgggcacctgcagc 7580

965 agcaggcgccagctgcctttcggcactactccctgcagctccagtcagcagcacgcgcgc 1024
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7581 agcagcagcagcagcagcagcagcagcttccataaccaccaccacagcagcacgcagct 7640

1025 ccacgcgcgccttggcataccaccaagctccaaagcagaagctgcagaacgcacttgcct 1084

7641 tctatgaccaccacagcagcagcagcagcagcagcagcagcagcagcagcttccatgacc 7700

1085 ccctctgccttccccagggtaccacttctcagcatttcccagcttcttccccacct 1144
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7701 accaccaggagaagcagcagcttccatgaccaccacagcagcaacagcagttccatgacc 7760

1145 cctccactactctctctctgtccagcttgcagcttggggcagggggcccaactctataagagtt 1204
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7761 accaccaccaccagcagcagcagcttccatgaccaccacagcagcaacagcagcttcc 7820

1205 gcacagcagcagctgccaccgcccatgacagcgcctgactgcagctccagctggccc 1264

7821 atgaccaccagcagcagcagcagcagcagcagcagcagcagcagcttccatgaccaccacgc 7880

1365 cgggcagcgggtccagaattctctctaccagctggccgcgcctcagctatgaccagc 1324
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7881 agaagcagcagcttccatgaccaccaccacccacccacagcagcagcttccatgaccac 7940

1325 agcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1384
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7941 agcagcagcagcagcagcttccatgaccaccacagcagcagcagcttccatgaccacc 8000

1385 ccagggaacccctccattacaaaacctcgcaagtatcgactacgactcagggcagcagggcc 1444
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8001 ccagcagaagcagcagcttccatgaccaccgccagcagcagcagcagcttccatgaccacc 8060

1445 agggctactgccagcgggagcagcagcttccgaccaccacagcagcagcttaccagaccttc 1503
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8061 accaccagcagcagcagcagcagcagcttccatgaccaccacagcagcagcagcagcttc 8119
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
AI22780
AA122780 standard; DNA: 1635 BP.

```

RESULT	12	
AAI22780		
AAI22780 standard; DNA; 1635 BP.		
AAI22780;		
12-OCT-2001 (first entry)		
Probe #12713 for gene expression analysis in human cervical cell sample.		
Probe; human; microarray; gene expression; cervical epithelial cell;		
Cervical cancer; ss.		
Homo sapiens.		
WC200157278-A2.		
09-AUG-2001.		
30-JAN-2001; 2001WO-US00670.		
04-FEB-2000; 2000US-0180312.		
26-MAY-2000; 2000US-0207456.		
30-JUN-2000; 2000US-0608408.		
03-AUG-2000; 2000US-0632366.		
21-SEP-2000; 2000US-0234887.		
27-SEP-2000; 2000US-0236359.		
04-OCT-2000; 2000GB-0024263.		

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-48901/53.

Human genome-derived single exon nucleic acid probes useful for  
analyzing gene expression in human cervical epithelial cells -

Claim 25; SEQ ID NO 12713; 487pp; English.

The present invention relates to human single exon nucleic acid probes  
(SENP). The present sequence is one such probe. The SENPs are derived  
from human HeLa cells. The SENPs can be used to produce a single exon  
microarray, which can be used for measuring human gene expression in a  
sample derived from human cervical epithelial cells. By measuring gene  
expression, the probes are therefore useful in grading and/or staging  
of diseases of the cervix, notably cervical cancer.

Note: The sequence data for this patent did not form part of the printed  
specification, but was obtained in electronic format directly from WIPO  
at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).

Sequence 1635 BP; 538 A; 916 C; 18 G; 163 T; 0 other;

Query Match 1.2%; Score 69.4; DB 22; Length 1635;  
Best Local Similarity 46.4%; Pred. No. 0.00024;  
Matches 332; Conservative 0; Mismatches 376; Indels 7; Gaps 3;

993 tcacttcctgcagctccagcagcacgcccggcccgacgccccctgtgatccaccagct 1052  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
675 tcactgccaccacacacacacacacacattactaccaccaccattaccaccacca 734  
1053 ccaaggcagaagctgcgaagacaattgctccctctgtcccttcccccagggtacc 1112  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
735 ccatactaccaccacacacacacacacattactaccaccaccaccaccattact 794  
1113 ctcttcagattccagctctccacacctctccactactctctctctgtccagg 1172  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
795 ccaccaccaccaccacca--cacaccattactaccaccaccaccaccaccattcac 852  
1179 tggtyggggggggggccactctctataagagtgcagcagcacgactgcccagct 1232  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
853 tacaccaccaccaccaccattactaccaccaccaccaccattactaccaccaccac 912  
1233 caggccgctgactgcagctccagctcagctggccccggggcagggggtccagattctc 1292  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
913 caccaccattactaccaccacc---accattaccattactaccaccattaccacca 968  
1293 ctaccagtgccgctcagctatgacacagcagcagcagcagcagcagcagcagcag 1352  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
969 ccaccaccattaccaccaccaccattaccattactaccaccaccattaccaccacca 1028  
1353 gcagcagcagcctttcagcgcgcacatgccacggcaggaacccctccattaccacaac 1412  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1029 caacattactaccaccaccaccaccaccaccaccattactaccaccaccaccacca 1088  
1413 gccaagtatccgactacgagcagcagcagcagcagcagcagcagcagcagcagc 1472  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1089 tcactaccaccaccaccaccattactaccaccaccaccaccattactaccaccacca 1148  
1473 ccgagcccccagcagctactacca--gacttcagccccagctccaccgcactccaccgc 1531  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1149 ccacattaccattactaccattaccaccaccaccaccaccattaccaccaccaccat 1208  
1532 gtccgtggggcgctcaocttctcagttccaccacgctgcgcgttgtgtccaaactgg 1591  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1209 ccatactaccaccaccaccaccaccattactaccaccattaccaccattaccaccacca 1268  
1592 agaacttccctacagccagcagcctcagcagcagcgggcttcccgcagggatactg 1651  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1269 tcactaacaccaccattactaccattaccaccattaccaccattactaccaccacca 1328  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||



Claim 25; SEQ ID No 8445; 322pp; English.

The present invention relates to novel single exon nucleic acid probes. The present sequence is one such probe. The probes are useful for measuring human gene expression in a human breast sample, where the probe hybridizes at high stringency to a nucleic acid expressed in the human breast. The probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human breast, particularly those diseases with polygenic aetiology. The diseases include: breast cancer, disorders of development, inflammatory diseases of the breast, fibrocystic changes, proliferative breast disease and non-carcinoma tumours.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 1635 BP; 538 A; 916 C; 18 G; 163 T; 0 other;

Query Match 1.2%; Score 69.4; DB 22; Length 1635;  
Best Local Similarity 46.4%; Pred. No. 0.00024;  
Matches 332; Conservative 0; Mismatches 376; Indels 7; Gaps 3;

993 tcactccctgacgtccagcagccacccgcccgcagcagccctggcgtaccacgaagt 1052  
1113 cttctccagattccagctctccacacccctccacacccctccctctgtccaggg 1172  
795 ccaccaccatcacacca--ccaccaccatcacaccacacacacacacacacacac 852  
1173 tgggtggcagggggccactccataagagttgcagcagcagcagcagcagcagcagc 1232  
853 tacaccaccaccaccaccaccaccaccaccaccaccaccaccaccaccaccaccacc 912  
1233 caggccgctgactgcagcgtccagctgctggccggggcagcgggtccagatcttcagc 1292  
913 caccaccaccatcacaccaccacc---accatcacatcacaccaccaccaccaccacc 968  
1293 ctaccagtcggccgctcagctatcacagcagcagcagcagcagcagcagcagcagc 1352  
969 ccaccaccatcacaccaccaccaccaccaccaccaccaccaccaccaccaccaccacc 1028  
1353 gcagcagcagcgtccagcggccagcagcagcagcagcagcagcagcagcagcagc 1412  
1029 ccaccatcacaccaccaccaccaccaccaccaccaccaccaccaccaccaccaccacc 1088  
1413 cgcacagatcacgactacggggcagcagcagcagcagcagcagcagcagcagcagcgt 1472  
1089 tcactaccaccaccaccaccaccaccaccaccaccaccaccaccaccaccaccaccacc 1148  
1473 ccgagccaccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1531  
1149 ccaccatcacaccaccaccaccaccaccaccaccaccaccaccaccaccaccaccacc 1208  
1532 gtcgtggcggcgtccactccatccagtcagtcagtcagtcagtcagtcagtcagtcag 1591  
1209 ccactaccaccaccaccaccaccaccaccaccaccaccaccaccaccaccaccaccacc 1268  
1592 agaaccttcctacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1651  
1269 tcactaacaccaccaccaccaccaccaccaccaccaccaccaccaccaccaccaccacc 1328  
1652 accagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1706  
1329 ccaccatcacaccaccaccaccaccaccaccaccaccaccaccaccaccaccaccacc 1383

RESULT 15

AA113556  
ID AA113556 standard; DNA; 1973 BP.  
XX AC AA113556;  
XX 12-OCT-2001 (first entry)  
XX DE Probe #3489 for gene expression analysis in human cervical cell sample.  
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;  
XX KW cervical cancer; ss.  
XX OS Homo sapiens.  
XX WO200157278-A2.  
XX 09-AUG-2001.  
XX 30-JAN-2001; 2001WO-US00670.  
XX 04-FEB-2000; 2000US-0180312.  
XX 26-MAY-2000; 2000US-0207456.  
XX 30-JUN-2000; 2000US-0608408.  
XX 03-AUG-2000; 2000US-0632356.  
XX 21-SEP-2000; 2000US-0234687.  
XX 27-SEP-2000; 2000US-0236359.  
XX 04-OCT-2000; 2000GB-0024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488901/53.  
XX Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human cervical epithelial cells -  
XX Claim 25; SEQ ID No 3489; 487pp; English.  
XX The present invention relates to human single exon nucleic acid probes  
XX (SENP). The present sequence is one such probe. The SENPs are derived  
XX from human HeLa cells. The SENPs can be used to produce a single exon  
XX microarray, which can be used for measuring human gene expression in a  
XX sample derived from human cervical epithelial cells. By measuring gene  
XX expression, the probes are therefore useful in grading and/or staging  
XX of diseases of the cervix, notably cervical cancer.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

Query Match 1.2%; Score 69.4; DB 22; Length 1973;  
Best Local Similarity 46.4%; Pred. No. 0.00025;  
Matches 332; Conservative 0; Mismatches 376; Indels 7; Gaps 3;

993 tcactccctgacgtccagcagccacccgcccgcagcagccctggcgtaccacgaagt 1052  
829 tcactgccaccaccaccaccaccaccaccaccaccaccaccaccaccaccaccaccacc 888  
1053 ccaaggcagagtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1112  
889 ccactaccaccaccaccaccaccaccaccaccaccaccaccaccaccaccaccaccacc 948  
1113 cttctccagattccagctctccacacccctccacacccctccctctgtccaggg 1172  
949 ccaccaccatcacacca--ccaccaccatcacaccaccaccaccaccaccaccaccacc 1006  
1173 tgggtggcagggggccactccctataagagttgcagcagcagcagcagcagcagcagc 1232  
1007 taccaccaccaccaccaccaccaccaccaccaccaccaccaccaccaccaccaccacc 1066





502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100 1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120 1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160 1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180 1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200 1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220 1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240 1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260 1261 1262 1263 1264 1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280 1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300 1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320 1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340 1341 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360 1361 1362 1363 1364 1365 1366 1367 1368 1369 1370 1371 1372 1373 1374 1375 1376 1377 1378 1379 1380 1381 1382 1383 1384 1385 1386 1387 1388 1389 1390 1391 1392 1393 1394 1395 1396 1397 1398 1399 1400 1401 1402 1403 1404 1405 1406 1407 1408 1409 1410 1411 1412 1413 1414 1415 1416 1417 1418 1419 14

FILING DATE: 26-AUG-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BENT, Stephen A.  
 REGISTRATION NUMBER: 29,783  
 REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703)836-9300  
 TELEFAX: (703)683-4109  
 TELEX: 899149  
 INFORMATION FOR SEQ ID NO: 14:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 7218 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 IMMEDIATE SOURCE:  
 CLONE: pTgpt-Fls  
 US-08-232-463-14

Query Match 1.1%; Score 63.8; DB 1; Length 7218;  
 Best Local Similarity 4.88; Pred. No. 0.00059;  
 Conservative 209; Mismatches 131; Indels 0; Caps 0;

```

Query Match      1.1%; Score 63.8; DB 1; Length 7218;
Best Local Similarity 4.8%; Pred. No. 0.00059;
Matches 17; Conservative 209; Mismatches 131; Indels 0; Gaps
0;
QY 5269 ttctaccctctgtgcggtgagagagacgcttcaccacacatgctacactgtctc 5328
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1031 TTCGAGCTGGCTCGAGTGCAGGAGCTTGCATATTTTTTTTTTTTTTTTTTTTT 1990
QY 5329 aactccctcgagatgagcccaagccacaggaagcctctctctctctctctctcc 5388
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1091 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1150
QY 5389 tcattctctctctctctctctgttgatgagcgggctccctgcgcacactcoctgga 5448
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1151 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1210
QY 5449 ggcctccatctctgagccgcggcctctctgcctctctctccacatgctctggggcct 5508
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1211 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1270
QY 5509 gtgtttccaagggcctgagtaacctctgtctgtttgtgctctgcgaaccggcgc 5568
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1271 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1330
QY 5569 aactccaaggacctgggggaacctgtgggcctactacactggaacctgctccccc 5625
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1281 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1387

```

RESULT 144  
US-08-209-747-1/c  
Sequence 1, Application US/08209747  
Patent No. 5733771  
GENERAL INFORMATION:  
APPLICANT: Lewis, Randolph V.  
TITLE OF INVENTION: cdnas Encoding Minor Ampullate Spider  
TITLE OF INVENTION: Silk Proteins  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-3487  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/209,747

FILING DATE: 14-MAR-1994  
 CLASSIFICATION: 530  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Murphy Jr., Gerald M.  
 REGISTRATION NUMBER: 28,977  
 REFERENCE/DOCKET NUMBER: 1447-104P  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 703-205-8000  
 TELEFAX: 703-205-8050  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2793 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: CDNA  
 HYPOTHETICAL: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Nephila clavipes  
 TISSUE TYPE: minor ampullate gland  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 183..2675  
 OTHER INFORMATION: /product="N. clavipes minor  
 OTHER INFORMATION: ampullate silk protein"  
 US-08-209-747-1

Query Match 1.18; Score 63.6; DB 1; Length 2793;  
 Best Local Similarity 51.0%; Pred. No. 0.0005;  
 Matches 203; Conservative 0; Mismatches 189; Indels 6; Gaps 2;  
 QY 1269 gaagcggtcccaaatcttcatgctcaccagtcgagcgcgctcgaagttgagccagagagga 1328  
 DB 1638 GCACTGCCCCAGACAGCGGACACGCTCCAGACACAGACCTCTTACCGTAACCCAGCG 1579  
 QY 1329 gaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1388  
 DB 1578 CCTCGGCTCCTGTGTCAGAGCTGTCAGAGACACAGACCTCTGCTCCGACACCTATCCA 1519  
 QY 1389 ggaacccctccattaccaaaacctgcagcaagatagagactagagcgagcagcagcagcag 1448  
 DB 1518 CCTTACACACGTAAGCTCTCCAGCTCTGCGCTGCGACAGCTCCAGCTCCAGCTCCAGCA 1459  
 QY 1449 ctactgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1508  
 DB 1458 CCTTACACGTAAGCTCTCCAGCTCTGCGCTGCGACAGCTCCAGCTCCAGCTCCAGCA 1404  
 QY 1509 cagctccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1568  
 DB 1403 CAGCTCTGTCAGACGCGACGACGTAACCTTGAACCTTGAACCTTGAACCTTGAACCTTGA 1345  
 QY 1569 gtcgcgcgtatgcaaaacctggaagacttccctacagcagcagcagcagcagcagcagcag 1628  
 DB 1344 GAGAGGAGCTCTGCTCTCCAGACCGGACGAGGAGCTCTGCTCTCCAGACACACACCTCA 1285  
 QY 1629 ggccttcccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1666  
 DB 1284 AACCAACAGGCGCTCCAGCTCTGCTCCAGACGCTGCA 1247

RESULT 5  
 US-08-458-298-1/c  
 Sequence 1, Application US/08458298  
 Patent No. 5756677  
 GENERAL INFORMATION:  
 APPLICANT: Lewis, Randolph V.  
 APPLICANT: Colglin, Mark  
 TITLE OF INVENTION: CDNA's Encoding Minor Ampullate Spider  
 TITLE OF INVENTION: Silk Proteins  
 NUMBER OF SEQUENCES: 56  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: Birch, Stewart, Kolasch & Birch

STREET: P.O. Box 747  
 CITY: Falls Church  
 STATE: Virginia  
 COUNTRY: USA  
 ZIP: 22040-3487  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/458,298  
 FILING DATE: 02-JUN-1995  
 CLASSIFICATION: 530  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/209,747  
 FILING DATE: 14-MAR-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Murphy Jr., Gerald M.  
 REGISTRATION NUMBER: 28,977  
 REFERENCE/DOCKET NUMBER: 1447-104P  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 703-205-8000  
 TELEFAX: 703-205-8050  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2793 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: CDNA  
 HYPOTHETICAL: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Nephila clavipes  
 TISSUE TYPE: minor ampullate gland  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 183..2675  
 OTHER INFORMATION: /product="N. clavipes minor  
 OTHER INFORMATION: ampullate silk protein"  
 US-08-458-298-1

Query Match 1.18; Score 63.6; DB 1; Length 2793;  
 Best Local Similarity 51.0%; Pred. No. 0.0005;  
 Matches 203; Conservative 0; Mismatches 189; Indels 6; Gaps 2;  
 QY 1269 gaagcggtcccaaatcttcatgctcaccagtcgagcgcgctcgaagttgagccagagagga 1328  
 DB 1638 GCACTGCCCCAGACAGCGGACACGCTCCAGACACAGACCTCTTACCGTAACCCAGCG 1579  
 QY 1329 gaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1388  
 DB 1578 CCTCGGCTCCTGTGTCAGAGCTGTCAGAGACACAGACCTCTGCTCCGACACCTATCCA 1519  
 QY 1389 ggaacccctccattaccaaaacctgcagcaagatagagactagagcgagcagcagcagcag 1448  
 DB 1518 CCTTACACACGTAAGCTCTCCAGCTCTGCGCTGCGACAGCTCCAGCTCCAGCTCCAGCA 1459  
 QY 1449 ctactgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1508  
 DB 1458 CCTTACACGTAAGCTCTCCAGCTCTGCGCTGCGACAGCTCCAGCTCCAGCTCCAGCA 1404  
 QY 1509 cagctccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1568  
 DB 1403 CAGCTCTGTCAGACGCGACGACGTAACCTTGAACCTTGAACCTTGAACCTTGAACCTTGA 1345  
 QY 1569 gtcgcgcgtatgcaaaacctggaagacttccctacagcagcagcagcagcagcagcagcag 1628  
 DB 1344 GAGAGGAGCTCTGCTCTCCAGACCGGACGAGGAGCTCTGCTCTCCAGACACACACCTCA 1285  
 QY 1629 ggccttcccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1666

Db 1284 AACACAGCGCCCTCAGCTCTGCTCCAGCAGCTGCA 1247

RESULT 6  
US-08-469-802B-3  
; Sequence 3, Application US/08469802B  
; Patent No. 5741645  
; GENERAL INFORMATION:  
; APPLICANT: Orr, Harry T.  
; APPLICANT: Rannum, Laura P.W.  
; APPLICANT: Chung, Ming-yi  
; APPLICANT: Zoghbi, Huda Y.  
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia

; Patent No. 5741645  
; TITLE OF INVENTION: Type 1 and Method for Diagnosis  
; NUMBER OF SEQUENCES: 47  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Muetting, Raasch, Gebhardt & Schwappach, P.A.  
; STREET: 119 No. 5741645th Fourth Street, Suite 203  
; CITY: Minneapolis  
; STATE: MN  
; COUNTRY: USA  
; ZIP: 55401

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/469,802B  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Muetting, Ann M.  
; REGISTRATION NUMBER: 33,977  
; REFERENCE/DOCKET NUMBER: 110.00030101  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 612-305-1217  
; TELEFAX: 612-305-1225  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 234 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA

US-08-469-802B-3  
Query Match 1.0%; Score 61.4; DB 1; Length 234;  
Best Local Similarity 54.7%; Pred. No. 0.00072;  
Matches 122; Conservative 0; Mismatches 101; Indels 0; Gaps 0;  
QY 1264 ccggggcagcgggtccagaattcttcagtcggtcgccgctcagctatgaccag 1323  
Db 7 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 66  
QY 1324 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1383  
Db 67 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 126  
QY 1384 gccaggaacaccttcattccaaacctcgcagctacgactacgagcagcagcag 1443  
Db 127 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 186  
QY 1444 cagggctactgcagcggcagcagcagcagcagcagcagcagcagcagcagcag 1486  
Db 187 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 229

US-08-469-802B-3  
Query Match 1.0%; Score 61.4; DB 1; Length 234;  
Best Local Similarity 54.7%; Pred. No. 0.00072;  
Matches 122; Conservative 0; Mismatches 101; Indels 0; Gaps 0;  
QY 1264 ccggggcagcgggtccagaattcttcagtcggtcgccgctcagctatgaccag 1323  
Db 7 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 66  
QY 1324 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1383  
Db 67 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 126  
QY 1384 gccaggaacaccttcattccaaacctcgcagctacgactacgagcagcagcag 1443  
Db 127 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 186  
QY 1444 cagggctactgcagcggcagcagcagcagcagcagcagcagcagcagcagcag 1486  
Db 187 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 229

RESULT 7  
US-08-267-803B-3  
; Sequence 3, Application US/08267803B

; Patent No. 5834183  
; GENERAL INFORMATION:  
; APPLICANT: Orr, Harry T.  
; APPLICANT: Rannum, Laura P.W.  
; APPLICANT: Chung, Ming-yi  
; APPLICANT: Zoghbi, Huda Y.  
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia  
; Patent No. 5834183  
; TITLE OF INVENTION: Type 1 and Method for Diagnosis  
; NUMBER OF SEQUENCES: 85  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Muetting, Raasch, Gebhardt & Schwappach, P.A.  
; STREET: P.O. Box 581415  
; CITY: Minneapolis  
; STATE: MN  
; COUNTRY: USA  
; ZIP: 55458-1415  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/267,803B  
; FILING DATE: 28-JUN-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McCormack, Myra H.  
; REGISTRATION NUMBER: 36,602  
; REFERENCE/DOCKET NUMBER: 110.00030120  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 612-305-1217  
; TELEFAX: 612-305-1228  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 234 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA

US-08-267-803B-3  
Query Match 1.0%; Score 61.4; DB 2; Length 234;  
Best Local Similarity 54.7%; Pred. No. 0.00072;  
Matches 122; Conservative 0; Mismatches 101; Indels 0; Gaps 0;  
QY 1264 ccggggcagcgggtccagaattcttcagtcggtcgccgctcagctatgaccag 1323  
Db 7 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 66  
QY 1324 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1383  
Db 67 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 126  
QY 1384 gccaggaacaccttcattccaaacctcgcagctacgactacgagcagcagcag 1443  
Db 127 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 186  
QY 1444 cagggctactgcagcggcagcagcagcagcagcagcagcagcagcagcagcag 1486  
Db 187 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 229

US-08-267-803B-3  
Query Match 1.0%; Score 61.4; DB 2; Length 234;  
Best Local Similarity 54.7%; Pred. No. 0.00072;  
Matches 122; Conservative 0; Mismatches 101; Indels 0; Gaps 0;  
QY 1264 ccggggcagcgggtccagaattcttcagtcggtcgccgctcagctatgaccag 1323  
Db 7 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 66  
QY 1324 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1383  
Db 67 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 126  
QY 1384 gccaggaacaccttcattccaaacctcgcagctacgactacgagcagcagcag 1443  
Db 127 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 186  
QY 1444 cagggctactgcagcggcagcagcagcagcagcagcagcagcagcagcagcag 1486  
Db 187 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 229

RESULT 8  
US-08-232-463-14/C  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS



FILING DATE: 07-NOV-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Petithory, Joanne R  
REGISTRATION NUMBER: 42,995  
REFERENCE/DOCKET NUMBER: 8600-0177.30  
TELEPHONE: 650-324-0880  
TELEFAX: 650-324-0960  
TELEX:

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3254 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
ORIGINAL SOURCE: Drosophila spry cdna  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 1323...3095  
OTHER INFORMATION:  
US-08-965-903B-1

Query Match 1.0%; Score 58.2; DB 3; Length 3254;  
Best Local Similarity 59.3%; Pred. No. 0,0062;  
Matches 99; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 1213 ccgaactgccagcccatgacagcgcctgtatgcagtcgacctggccccggcgag 1272  
DB 1365 CCGGCTGCCAGCAGGATTAATGGCAGACAGCCACCACCTGTGTGTCGCCCGC 2024  
QY 1273 cgggttcgaattttatcctaccagtgcggccctcagctatgaccagcagcag 1332  
DB 2025 CGGAATPCAGCACTGCACCTGCAACAACACACAGCACACCTGCAAGCAGCAG 2084  
QY 1333 cagcacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1379  
DB 2085 CAGCACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 2131

RESULT 11  
US-08-469-802B-2  
Sequence 2, Application US/08469802B  
Patent No. 5741645  
GENERAL INFORMATION:  
APPLICANT: Orr, Harry T.  
APPLICANT: Ranum, Laura P.W.  
APPLICANT: Chung, Ming-vi  
APPLICANT: Zoghbi, Huda Y.  
TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia  
Patent No. 5741645  
TITLE OF INVENTION: Type 1 and Method for Diagnosis  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Muetting, Raasch, Gebhardt & Schwappach, P.A.  
STREET: 119 N. 5741645th Fourth Street, Suite 203  
CITY: Minneapolis  
STATE: MN  
COUNTRY: USA  
ZIP: 55401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,802B  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Muetting, Ann M.

FILING DATE: 07-NOV-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Petithory, Joanne R  
REGISTRATION NUMBER: 42,995  
REFERENCE/DOCKET NUMBER: 8600-0177.30  
TELEPHONE: 650-324-0880  
TELEFAX: 650-324-0960  
TELEX:

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3254 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
ORIGINAL SOURCE: Drosophila spry cdna  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 1323...3095  
OTHER INFORMATION:  
US-08-965-903B-1

Query Match 1.0%; Score 58.2; DB 3; Length 3254;  
Best Local Similarity 59.3%; Pred. No. 0,0062;  
Matches 99; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 1213 ccgaactgccagcccatgacagcgcctgtatgcagtcgacctggccccggcgag 1272  
DB 1365 CCGGCTGCCAGCAGGATTAATGGCAGACAGCCACCACCTGTGTGTCGCCCGC 2024  
QY 1273 cgggttcgaattttatcctaccagtgcggccctcagctatgaccagcagcag 1332  
DB 2025 CGGAATPCAGCACTGCACCTGCAACAACACACAGCACACCTGCAAGCAGCAG 2084  
QY 1333 cagcacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1379  
DB 2085 CAGCACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 2131

RESULT 11  
US-08-469-802B-2  
Sequence 2, Application US/08469802B  
Patent No. 5741645  
GENERAL INFORMATION:  
APPLICANT: Orr, Harry T.  
APPLICANT: Ranum, Laura P.W.  
APPLICANT: Chung, Ming-vi  
APPLICANT: Zoghbi, Huda Y.  
TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia  
Patent No. 5741645  
TITLE OF INVENTION: Type 1 and Method for Diagnosis  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Muetting, Raasch, Gebhardt & Schwappach, P.A.  
STREET: 119 N. 5741645th Fourth Street, Suite 203  
CITY: Minneapolis  
STATE: MN  
COUNTRY: USA  
ZIP: 55401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,802B  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Muetting, Ann M.

[illegible]

D6		GT	1437	gcaagcgcaaggctactgcacgcgga	1463
D8		121	ggagcagcgcagcacacacacacaca	147	

RESULT 14  
US-08-267-803B-6  
; Sequence 6, Application US/08267803B  
; Patent No. 5834183  
; GENERAL INFORMATION:  
; APPLICANT: ORT, HARRY T.

APPLICANT: Ranum, Laura P.W.  
APPLICANT: Chung, Ming-yi  
APPLICANT: Zoghbi, Huda Y.  
TITLE OF INVENTION: Gene Sequence for Sphocerebellar Ataxia

TITLE OF INVENTION: Type 1 and method for diagnosis  
 NUMBER OF SEQUENCES: 85  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Metling, Raasch, Gebhardt & Schwappach, P.A.  
 STREET: P.O. Box 581415  
 CITY: Dallas, Texas 75215-1415  
 COUNTRY: U.S.A.

1 CITY: Minneapolis  
2  
3 STATE: MN  
4  
5 COUNTRY: USA  
6  
7 ZIP: 55458-1415  
8  
9 COMPUTER READABLE FORM:  
10  
11 MEDIUM TYPE: floppy disk  
12  
13 COMPUTER: IBM PC compatible  
14  
15 OPERATING SYSTEM: PC-DOS/MS-DOS  
16  
17 SOFTWARE: PatentIn Release #1.0, Version #1.25  
18  
19 CURRENT APPLICATION DATA:  
20  
21 APPLICATION NUMBER: US/08/267,803B  
22  
23 FILING DATE: 28-JUN-1994  
24  
25 CLASSIFICATION: 435

1 CLASSIFICATION INFORMATION:  
2 ATTORNEY/AGENT INFORMATION:  
3 NAME: MCCORMACK, MYRA H.  
4 REGISTRATION NUMBER: 36,602  
5 REFERENCE/DOCKET NUMBER: 110,00030122  
6 TELECOMMUNICATION INFORMATION:  
7 TELEPHONE: 612-305-1217  
8 TELEFAX: 612-305-1228  
9 INFORMATION FOR SEQ ID NO: 6:  
10 SEQUENCE CHARACTERISTICS:

```

? LENGTH: 154 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA
? US-08-267-803B-6

```

	Query Match	1.08;	Score 57.4;	DB 2;	Length 154;
	Best Local Similarity	61.9%;	Pred. No. 0.0041;		
	Matches	91;	Mismatches	56;	Gaps
QY	1317	tgaccagcagcacgcacgacgacgacgacgacgacgaagaaccttcaggcg	137		
DG	1	TGAGAGAGAGACACACACACACACACACACACACACACACACACAGCA	60		
QY	1377	gcacactgccggcgaagaacctccattaccaaacctctgcgaagtatcagactcaaggcca	143		

D6 61 GCGACGACGCAGAGCAAGCACAAGCAGCACAGCAGCGACGACGAGA 120

D7 1437 gcaagggccagggtctactccagcccgga 1463

D8 121 GCATCATCCGACGAGCAGCAGCAGCT 147

[illegible]

Search completed: December 4, 2001, 16:11:33  
Job time: 7056 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2001, 14:52:14 ; Search time 2862.31 seconds  
(without alignments)  
22607.991 Million cell updates/sec

Title: US-09-508-821B-5

Perfect score: 6022

Sequence: 1 ggaaccagcagcccaaggg.....ggaagctcttggtgctgcag 6022

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapept 1.0

Searched: 11351937 seqs, 5372899281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_hic:\*  
10: gp\_est1:\*  
11: gp\_est2:\*  
12: gp\_hic:\*  
13: gp\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_tod:\*  
20: em\_gss\_vit:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	681.6	9.6	865	11	BE869510 601446178
2	577.6	9.6	724	11	BG828494 602751486
3	575.6	9.6	1012	11	BF307287 601891563
4	546.2	9.1	746	10	AL119277 DKFZP61A
5	527.2	8.8	566	10	AM860154 RCO-CT037
6	527.2	8.8	566	10	AM860154 RCO-CT037
7	517.8	8.6	963	10	AI893920 mc85a07.Y
8	513.6	8.5	558	10	AM860153 RCO-CT037
9	513.6	8.5	558	10	AM860153 RCO-CT037
10	507.8	8.4	773	11	BG422462 602449035
11	506	8.4	540	10	AM821875 RCO-ST037
12	506	8.4	540	10	AM821899 RCO-ST037

13	424	7.0	832	11	BF304773
14	390.6	6.5	501	13	AA642061
15	373.4	6.2	613	10	AA653431
16	361.2	6.0	940	11	BF234313
17	357.6	5.9	677	10	AI385630
18	315.4	5.2	611	10	AA052617
19	311	5.2	841	11	BF337555
20	285.6	4.7	504	10	BB653219
21	245	4.1	269	11	BF360615
22	240.4	4.0	274	11	BI000630
23	237.4	3.9	239	11	BF094022
24	231.8	3.8	1045	11	BF527692
25	230.4	3.8	232	10	AM062659
26	228	3.8	236	10	AM062658
27	214.4	3.6	570	11	BI259831
28	213	3.5	1634	12	AK013909
29	209.4	3.5	803	11	BG976042
30	207.6	3.4	675	10	AI547182
31	182.6	3.0	750	10	AI547124
32	175.4	2.9	212	10	AM176779
33	168.8	2.8	223	10	AM845544
34	164.4	2.7	158	10	AM492162
35	156	2.6	224	10	BE651204
36	153.2	2.5	452	10	BE928740
37	149.6	2.5	157	11	BF342418
38	148	2.5	186	11	BF734241
39	130.2	2.2	192	11	BF734241
40	128.6	2.1	146	10	BE169703
41	123	2.0	542	10	BE169703
42	116.4	1.9	146	10	BF872642
43	112.4	1.9	114	11	AQ741044
44	108.2	1.8	822	13	BF928778
45	106.4	1.8	108	11	BF928778

#### ALIGNMENTS

RESULT 1  
BE869510  
LOCUS 601446178F1 NIH\_MGC\_65 Homo sapiens CDNA clone IMAGE:3850146 5',  
DEFINITION mRNA sequence.  
ACCESSION BE869510.1 GI:10318286  
VERSION EST.  
KEYWORDS human.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (Bases 1 to 865)  
NIH-MGC http://mgc.ncl.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
CONTACT: Robert Strausberg, Ph.D.  
COMMENT Email: cgapbs-rt@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA library Preparation: Life Technologies, Inc.  
CDNA library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
http://llmg568.row:1 column: 19  
High quality sequence stop: 662.  
Location/Qualifiers  
1..865  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3850146"  
/clone.lib="NIH\_MGC\_65"  
/tissue\_type="adenocarcinoma"  
/lab\_host="DH10B (phage-resistant)"



Db	383	GGCCGCGCTTCCTCCGTGCAACGCGCTCCAGGACAGAGAGCCCTTACCAGGCGCATGCTG	442
QY	830	gtgagagaaagccttcaagcttgggaggccccacacccaccccccacaaagccagagagccac	889
Db	443	GTGAAGAGAGACCTCTACAGGTTGTGGGGGGGCCCCACACCCACACCCACACCCGCGAGCCAC	502
QY	890	taccctgcaggggtggtgcacagatgatatgaactgtatgaanaagaaagacagatgccccca	949
Db	503	TACCTGGACAGGGGTGGCCAAAGTATGATGAGACCTGATGATATAAAAAGACAGAGTGCCTCCCA	562
QY	950	gcagagcagtatgtcaagagcagagggggcccaaggtgcctttcgagactaacctccc--gcacgt	1007
Db	563	GGAGGCAATATGAGAGCGAGGGGCCCCACAGTGTGCTTCCTTCGAGCTCACATCCTTGGACCGT	622
QY	1008	ccagagacgaacccgcgcgcacccacacgcgcctctgcatatacccacagctccaaagagcagaagct	106
Db	623	CCAGAGAGACACCGCG-CCCAAGACAGACCTCGGATA-CCCAAGTCCCAAGGSCAAGACT	680
QY	1068	gcagagagacatgctgctccctctgcgccttccccaagtgtaacc	1111
Db	681	GCAGAAACACATTCCTCCCTCCCTGCTACCTTCCCCAGGGGTACC	724

[illegible]

REFERENCE (bases 1 to 1012)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
[strausberg@mail.nih.gov](mailto:strausberg@mail.nih.gov)

cDNA library Preparation: ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: Mcgill clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov](http://image.llnl.gov)  
 Plate: L10CM104 row: m column: 16  
 High quality sequence stop: 736.

**FEATURES**  
**source**

source

1. 1012  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4137399"  
/clone\_id="NH\_MGC\_17"  
/tissue\_type="thadomyosarcoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: muscle; Vector: pOT57; Site\_1: EcoRI;  
Site\_2: XhoI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(C). Size-selected  
for average insert size 1.8kb. Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(stratagene) and Superscript II RT (Life Technologies)."

260 a 307 c 295 g 148 t 2 others

Query Match	9.6%;	Score 575.6;	DB 11;	Length 1012;
Best Local Similarity	96.0%;	Pred. No. 1.1e-90;		
Matches 632;	Conservative	0;	Mismatches 21;	Indels 5;
				Gaps 4

QY	472	agataacagcccggtctatgacggtcttttcgagaaagtggtttccatctgcaacaa	531
Db	84	AGATAACCGCCCGAGTCATGACGTCCTTTTGAAAAAGTGTGGTTTCATGGCAACAA	143
QY	533	cagaactacacagcagactctgcagagaaacatacagcctagagaaattacagcagccgagt	551
Db	144	CAGAACTACACAGCAGACTCTGCAGAAACATCAAGCCTAGAAATTTACAAGCAGCCGAGT	203
QY	592	caggccgggctaatgctgagacccgagcgggtgtgtctgccaagagagcttatatacccgag	651
Db	204	CAGGCCGGGCTAATGCTGAGACCCGAGCAGCGCGTGTGTCACCAAGAGACTTTATTAACCCGAG	263
QY	652	ccttaccgcgagctatgagaggtgtgctgtgcagacgcctctgagactgagccgggtggcc	711
Db	264	CCTTACCAGACTATGAGAGGTGTGCTGTGCAGACGCCTCTGAGACTGACGACGCGGGGTGCC	323
QY	712	gcgcgaactagatcacacacgagcagcagcgcctgtcccacacagcagaagcctgcagaggagg	771
Db	324	GCGCGAAGTACACACGAGGAGGAGGAGGCGCTGCCACACAGAGAGGCTGCGAGGGGAGAG	383
QY	772	ccgggttccctctgctacagcgtgtccagagacagccctacacccagcgcgtatgtgtgt	831
Db	384	CCGGTTCCTCTGCTACAGCGTGTCCAGAGACAGACCCCTAACCGAGGCGCTATGCTGTT	443
QY	832	gagagagacctttagcgttgggtgggccccacagccacacccccacagccgagcagcaacta	891
Db	444	GAGAGAGACCTTAGCGTTGGGGGGCCCCACAGACACACCCCAACGCCACGCCACG - CACTA	502
QY	892	cctctcagaggtgtgcacagatagatagagaaactgtatgaaaaagacagcaggtgcgccccagc	951
Db	503	CCCTCAGAGGTGTGCACAGATAGATAGAAACTGTGATGAAGAAAGACAGCAGATG - CCCCCAC	561
QY	952	agagcagtatgacagagcagagcgcacacagtgagccttctcgactactcctctgcagctccag	1011
Db	562	AGGCAATATGACAGAGCAGGCGGCCCAAGTGTGCTTGTGCACTACCTCCCTGACAGTCCAG	621
QY	1012	cagcacaacgcgcgccacagcagccctctgcatatcccaaggttccaagagcaaaagc - tga	1070
Db	622	CAGCCCAAGAGAGGCCACAGAGGCCCTGTGATTAACCAAGAGCTCCAAAGAGAGAACTTGA	681
QY	1071	gaagcagacttcttcccccctctgccttcccccaaggtatcccaattctctcagcttcc	1128
Db	682	GAAAGCAATAGCTCCCTCTGCTTCCCTCCCC - AAGGTACACACTTTCCACAGGATTC	737

RESULT	4				
AL119277					
LOCUS	AL119277	746 bp	mrna	EST	25-FEB-2000
DEFINITION	DEF2761A121.1	r1.761	(synonym: hairy2)	Homo sapiens	cdna clone
	DEF2761A121.5		mrna sequence.		

ACCESSION	AL119227	
VERSION	AL119277.1	GI:5925176
KEYWORDS	EST,	
SOURCE	human.	
ORGANISM	Homo sapiens	

REFERENCE  
AUTHORS  
1 (bases 1 to 746)  
Ottenwaelder, B., Obermaier, B., Mewes, H.W., Gassenhuber, J. and

TITLE  
 Es\* (Ottenswælder, et al.)  
 JOURNAL  
 Unpublished (1939)  
 COMMENT  
 Contact: Ottenswælder B

Mr's  
Am Klopferplatz 18a D-82155 Martinsried, Germany  
This is the 5' sequence of the clone insert  
clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de,  
sequenced by MedGenomix (Martinsried/Germany) within the CDNA  
sequencing consortium of the German Genome Project. No 51 sequence  
available.  
This clone (DKFZ761A121) is available at the RZPD in Berlin.  
Please contact the RZPD; Ressourcenzentrum, Heubnerweg 6, 14059



Query Match	8.8%;	Score 527.2;	DB 10;	Length 566
Best Local Similarity	98.9%;	Pred. No. 2.7e-82;		

Email: [mouseest@watson.wustl.edu](mailto:mouseest@watson.wustl.edu)  
 This notice is available royalty-free through JINT; contact the

This clone is available royalty-free through LINL; contact the IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.

2

```

/db_xref="taxon:9606"
/clone_lib="CT0379"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORSTS PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
98 a 207 c 147 g 106 t

```

Query Match	8.5%;	Score 513.6;	DB 10;	Length 558;
Best Local Similarity	98.6%;	Pred. No. 6.5e-80;		
Matches 550; Conservative	0;	Mismatches 4;	Indels 4;	Gaps 3;

QY 5274 acccttcgtgcggttbgagaaagcgcgcttcacacacatagtactgtgtcaactc 5333  
|||||  
Db 2 ACCCTTCGTCGGGGGAG - AGGGAGACGGTCACCAACCATATGCACCTGTGTCAACTC 60

[illegible][illegible]

OY		5453	cacatctgcagccgaggcccctccttggccccctctcttcacagatgatcttggtggccctgtg	5512
D8		181	CCATCTGCACACCAGGCCCTCCTTCGTGCCCCCTTCCACAGATGACATTGGGGGCGCTGTGG	240

[illegible][illegible]

**Dy** 5633 agccaaactcaggaagtgcggcacgaagcaccctgtgaggagcctcgctgccg 5692  
**Db** 361 AGCCAAACTCAGGAGAAGTGCGCCACGAAGCAACCCTGTAGGAGGGCTCCCTGCCGC 420

QY 5693 tttagaggaactcaaggtcccgagtgtgcagctgcccacatgccggagaagccccca 5752  
|||||  
Db 421 TTGAGAGACACTCAAGGTCCCGAGTGTGACGCTGCCGCACATGCCGGAGAAGCCCCCA 480

```
QY      5753 gg---tgaagcccaagtgcaccgccaagcagggcccactgtcgaccaaagtgcacggaggcc 5810
        ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      481 GGCCTGAGCGCCCACTGATTCGGGCCAACAAGAGGGGCCCACTGCAGACCAATTGCTTGGGGCC 540
```

QY	5811	tgccccgagagctgcaga	5828
Db	541	tgccccgagagctgcaga	558

RESULT 9  
AW86383

LOCUS	558 bp	nrNA	EST	15 MAY 2000
DEFINITION	RC0-CT0379-060200-032-b08 CT0379 Homo sapiens cDNA, mRNA sequence			
ACCESSION	AB862383			
VERSION	AB862383.1 GI:7958081			

KEYWORDS  
EST.  
SOURCE  
ORGANISM  
human.  
Homo sapiens  
Phylogeny: Metazoa: Chordata; Craniata; Vertebrata; Euteleostomi;  
Elkayvta: Metazoa: Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
AUTHORS

1 (bases 1 to 558)  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homin.  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Britones, M.R.  
Necai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,

TITLE	JOURNAL	COMMENT
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)	Contact: Simpson A.J.G. Laboratory of Cancer Genetics
Goldman, G.H., Carralho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brustein, N.I., deoliveira, P.S., Bucher, P., Jongejean, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.		

Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel.: +55-11-3706002

Fax: +55-11-2707001  
Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
This sequence was derived from the FAPESP/LIR Human Cancer Genome Database and can be seen in the following IPI:

(http://www.ludwig.org.br/scripts/gethtml2.p?tl=st2=NC0-CT0379-06  
200-032-p08et3=2000-02-06et4=1)  
Seq. primer: puc 18 forward

High quality sequence stop: 558.  
 FEATURES:  
 Location/Qualifiers  
 1..558  
 source

```

/db_xref="taxon:9606"
/clone_lib="CR0379"
/dev_stage="Adult"

```

Small; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 19,716 - Ludwig Institute for Cancer Research) profiles

BASE COUNT      98 a      207 c      147 g      106 t

ORIGIN	Query Match	8.5%	Score 513.6;	DB 10;	Length 558;
--------	-------------	------	--------------	--------	-------------

Best Local Similarity	50.0%	Reads: 10,000,000	50.0%
* Matches	550;	Conservative	0; Mismatches 4; Indels 4; Gaps 3
<p>5274 acccttcgtgcggtgtgagagcgcagacgcgtttcacccacatagcactgtttgcaactc 5333</p>			

D6 2 ACCCTTCGTCGGGTGGAG - AAGGAGACGCTTCACACACATATGCACGTTGCAACTC 60

5334 ccctg - gagatgcgcaccaagccccacacaggaaacctctctctctctctctctat 5392

Db 61 CCTGTGAGATGCGCCCAAGCCCCACAGAGCCCTTCCTCTCTCCCTCTCTCTCAT 120

Db 121 CCGCTCCCTGTTCTCCITGGATGACAGCGGGGCCCTCCCTGGCCACACCTCCCTGGAGGCT 180

Db 181 CCATCTCGACGCCGCGCCCTCTCTGCCCCCTCTCCACAGATGCACTTGGGCGCTGTGG 240

5573 tcaaggaaccttggaggaacctctgtgqgcctactaaccttgaaacttccctcccaaaaaga 5632

5632 *atccaaactcaagagaaagatcagccgacacctgtatgaagagccctgcctgcgcgc* 5692  
 Db 301 *TCAAGACCTTGGGGACCTCTGTGGCCCTACTACCTGAACACTGCTCTCCCAAAAGA* 360

569    ttacagaaacactcaaaagatccccaattgcaactgccacactgccacgaagccccc 572  
 Db    361    AGCCAAACtCAAGGAGAGGTGCGGCCAGAAAGCACCTGTAAAGAGGCTTCGTGCCG 420



DB 421 TTGAGAGAACTCAAGTCCGAGTGTGCAGCTGCCGACCTGGCGGAGGCCGCCCA 480  
5753 gg--tgacggccagctgacccggccagcagggccactgcgacacagtgccgggccc 5810  
DB 481 GGCTGACGGCCAGCTGATCGCGAAGCAGGGGCCACTCGGCACCACTGCTGGGCC 540  
QY 5811 tgcccgagagctcgaga 5828  
DB 541 TGTCCGGAGGCTGCAGA 558  
RESULT 10  
BG422462 773 bp mRNA EST 14-MAR-2001  
60244903F1 NIH\_MGC\_14 Homo sapiens cDNA clone IMAGE:4587632 5',  
mRNA sequence.  
BG422462  
BG422462.1 GI:13328968  
EST.  
SOURCE  
Homo sapiens  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 773)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-re@mail.nih.gov  
Tissue Procurement: DCTD/PTP  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LILNL at:  
<http://image.llnl.gov>  
Plate: LILNL1318 row: e column: 09  
High quality sequence stop: 657.  
Location/Qualifiers  
1. 773  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4587632"  
/clone\_lib="NIH\_MGC\_14"  
/clone\_type="renal cell adenocarcinoma"  
/lab\_host="Dh10B (phage-resistant)"  
/note="Organ: kidney; Vector: pOMB7; Site\_1: XhoI; Site\_2:  
EcoRI; CDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCACGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 158 a 239 c 253 g 123 t  
ORIGIN  
Query Match 8.4%; Score 507.8; DB 11; Length 773;  
Best Local Similarity 98.2%; Pred. No. 7e-79;  
Matches 545; Conservative 0; Mismatches 7; Indels 3; Gaps 3;  
QY 5469 gccctctgcctctctccacagtcacttgctggccctgtgttcccaaggccctgag 5528  
DB 6 GGCTCTCTGGCCCTCTCTCCACAGTGCATTTGGGGCCCTGTGGTTTCCAGGCCCTGAG 65  
QY 5529 taactctgctctgttctgctctgcgaacccggccacttcaaggaaccttgaggga 5588  
DB 66 TACCTCTTGCTTG-TTCTTGCTCTGCCAARACCGGCGACCTTCAGGACCTTGGGA 124  
QY 5589 cctctgaggccctactaccctgaacactgctctcccaaaagagccaaactcaagga 5648

DB 125 CCTCTGTGGCCCTACTACCTGAACACTGCTCCCAAAAGAGCCAAACTCAAGGA 184  
QY 5649 gaagggtggccagagaagcacctgtgagaggcctcgtgcgcttgagagacactcaa 5708  
DB 185 GAAGGTGGGCCAGAGGACCTGTGTGAGGAGGCTCGCTGCCGCTTGAGAGACTCAA 244  
QY 5709 aggtcccgagtgagctgcagctgcgacactgcgggaagcccccaggtgacggccagct 5767  
DB 245 AGTTCGCGAGTGTGAGCTGCCGCTCGCGGGAAGCCCCCGGAGGCTGACGCCAGCT 304  
QY 5768 gaccggcccaagcaggcccaactgcgacacagtgccgggggctgtcccggaggctgcag 5827  
DB 305 GACCGGCCCAACAGAGGCCCACTGCGCACCACTGCTGCCGGG-CTGTCCGGAGGCTGCAG 363  
QY 5828 agctgctactgtgtgagtcggccggaggatggggcgagagagagagagagagagag 5887  
DB 364 AGCTGCTACTGTGTGATGCGCGGAGGATGGGGCGGAGGAGGAGAGGAGGAGGAG 423  
QY 5888 ggtcgcaaacatgactgcagcaagagagctccgcgcagagcccgcgaggagagagagag 5947  
DB 424 GGTCCCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 483  
QY 5948 cactgggtgcatgagcctgtgcccgtgtgagcggcggtcctacctgtgagcggggaag 6007  
DB 484 CACTGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 543  
QY 6008 ctctttgggtgcag 6022  
DB 544 CTCCTTGGGCTGCAG 558  
RESULT 11  
AW821875 540 bp mRNA EST 17-MAY-2000  
RCO-ST0379-190100-032-d02 ST0379 Homo sapiens cDNA, mRNA sequence.  
AW821875  
AW821875.1 GI:7914869  
EST.  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 540)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,  
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
2020663  
COMMENT  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-RCO-ST0379-190>)  
100-032-d02st3-2000-01-19st4-1)  
Seq primer: puc 18 forward  
High quality sequence start: 20  
High quality sequence stop: 540.  
Location/Qualifiers  
1. 540  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="ST0379"  
FEATURES  
source



/dev\_stage="Adult"  
 /note="Organ: stomach; Vector: puc18; Site\_1: SmaI;  
 Site\_2: SmaI; A mini-library was made by cloning products  
 derived from ORESSES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the puc 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."  
 BASE COUNT 97 a 203 c 140 g 100 t  
 ORIGIN

Query Match 8.4%; Score 506; DB 10; Length 540;  
 Best Local Similarity 98.5%; Pred. No. 1.4e-78;  
 Matches 532; Conservative 0; Mismatches 5; Indels 3; Gaps 2;

QY 5291 aagaagcagcagcgttcaacacatgatgacgtgttaactcccttgagatgagccca 5350  
 |||||  
 Db 2 AGAAGGAGAGCGG-TCACACACATATGCACTGTGTCCACCTCCCTGAGATGCGGCCA 60  
 |||||  
 QY 5351 agccccacaggaagccttctctctctctctctctctctctctctctctctct 5410  
 |||||  
 Db 61 AGCCCCAGAGAGAGCCTTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 120  
 |||||  
 QY 5411 tagatgacgagcggcctctctctctctctctctctctctctctctctctctct 5470  
 |||||  
 Db 121 TGAATGACAGCGGGGCTCTCCTGCGCACACTCCCTGAGAGCTCATCTGACCGCGGC 180  
 |||||  
 QY 5471 cctctctgacct 5530  
 |||||  
 Db 181 CCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 240  
 |||||  
 QY 5531 cctctctgctgttctctctctctctctctctctctctctctctctctctct 5590  
 |||||  
 Db 241 CCTCTTGCCTTGTCT 300  
 |||||  
 QY 5591 tctctgagcctctactacacctgaacactgctctctctctctctctctctctct 5650  
 |||||  
 Db 301 TCTGTGCGCCCTACTACCTGACACTGCTCCTCCCAAAAAGCAAACTCAAGAGAG 360  
 |||||  
 QY 5651 aggtgagcagagaagcactgctgagagagcctgctgctgctgagagaactcaag 5710  
 |||||  
 Db 361 AGGTGCGCCAGAGAGCAGCTGTGAGAGAGCCTCCTCCTCTGTGAGAAACACTCAAG 420  
 |||||  
 QY 5711 gtccgagctgtgagctgctgagcactgctgagagagcctctgagagcctgag 5768  
 |||||  
 Db 421 GTCCGAGTGTGAGCTGTGCGGACACTGCGGAGAGCCCTGAGCCTGAGCCTGAG 480  
 |||||  
 QY 5769 accgagcagagcagcagcactgctgagcagcagcagcagcagcagcagcag 5828  
 |||||  
 Db 481 ATCCGCGCCAGAGAGGCGCCACTGCGCACAGTGTGCTGTGCGGAGCTGTGAGAG 540  
 |||||

## RESULT 12

AM821899 540 bp mRNA EST 17-MAY-2000  
 LOCUS  
 DEFINITION RCO-ST0379-210100-032-d02 ST0379 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION AM821899  
 VERSION AM821899.1 GI:7914893  
 KEYWORDS  
 EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

## REFERENCE

1 (bases 1 to 540)  
 Dias Neto, E., Garcia Correa, R., Verjovsky-Almeida, S., Briones, M.R.,  
 Nagai, M.A., da Silva, N. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H.,  
 Brunstein, A., de Oliveira, P.S., Bucher, P., Tongenel, C.V., O'Hare  
 M.J., Soares, F., Brentani, K.R., Reis, D.F., de Souza, S.J. and  
 Simpson, A.V.  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 MEDLINE 20202663  
 COMMENT Contact: Simpson A.J.G.

Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil

Tel: +55-11-2704922  
 Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?L=412-RCO-ST0379-210  
 100-032-d02&3-2000-01-21&4-1)

Seq primer: puc 18 forward

High quality sequence start: 20

High quality sequence stop: 540.

## FEATURES

source

1. 540

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_id="ST0379"

/dev\_stage="Adult"

/note="Organ: stomach; Vector: puc18; Site\_1: SmaI;  
 Site\_2: SmaI; A mini-library was made by cloning products  
 derived from ORESSES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the puc 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."

BASE COUNT 97 a 203 c 140 g 100 t  
 ORIGIN

Query Match 8.4%; Score 506; DB 10; Length 540;  
 Best Local Similarity 98.5%; Pred. No. 1.4e-78;  
 Matches 532; Conservative 0; Mismatches 5; Indels 3; Gaps 2;

QY 5291 aagaagcagcagcgttcaacacatgatgacgtgttaactcccttgagatgagccca 5350  
 |||||  
 Db 2 AGAAGGAGAGCGG-TCACACACATATGCACTGTGTCCACCTCCCTGAGATGCGGCCA 60  
 |||||  
 QY 5351 agccccacaggaagccttctctctctctctctctctctctctctctctctct 5410  
 |||||  
 Db 61 AGCCCCAGAGAGAGCCTTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 120  
 |||||  
 QY 5411 tagatgacgagcggcctctctctctctctctctctctctctctctctctctct 5470  
 |||||  
 Db 121 TGAATGACAGCGGGGCTCTCCTGCGCACACTCCCTGAGAGCTCATCTGACCGCGGC 180  
 |||||  
 QY 5471 cctctctgacct 5530  
 |||||  
 Db 181 CCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 240  
 |||||  
 QY 5531 cctctctgctgttctctctctctctctctctctctctctctctctctctct 5590  
 |||||  
 Db 241 CCTCTTGCCTTGTCT 300  
 |||||  
 QY 5591 tctctgagcctctactacacctgaacactgctctctctctctctctctctctctct 5650  
 |||||  
 Db 301 TCTGTGCGCCCTACTACCTGACACTGCTCCTCCCAAAAAGCAAACTCAAGAGAG 360  
 |||||  
 QY 5651 aggtgagcagagaagcactgctgagagagcctgctgctgctgagagaactcaag 5710  
 |||||  
 Db 361 AGGTGCGCCAGAGAGCAGCTGTGAGAGAGCCTCCTCCTCTGTGAGAAACACTCAAG 420  
 |||||  
 QY 5711 gtccgagctgtgagctgctgagcactgctgagagagcctctgagagcctgag 5768  
 |||||  
 Db 421 GTCCGAGTGTGAGCTGTGCGGACACTGCGGAGAGCCCTGAGCCTGAGCCTGAG 480  
 |||||  
 QY 5769 accgagcagagcagcagcactgctgagcagcagcagcagcagcagcagcag 5828  
 |||||  
 Db 481 ATCCGCGCCAGAGAGGCGCCACTGCGCACAGTGTGCTGTGCGGAGCTGTGAGAG 540  
 |||||

RESULT 13  
BF304773 832 bp mRNA 21-NOV-2000  
601888356F1 NH\_MGC\_17 Homo sapiens cDNA clone IMAGE:4122469 5',  
EST  
RNA sequence.  
BF304773  
BF304773.1 GI:11251671  
EST.  
SOURCE  
Homo sapiens  
human.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 832)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
cDNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov](http://image.llnl.gov)  
Plate: LHCMI005 row: 0 column: 14  
High quality sequence stop: 619.  
Location/Qualifiers  
1. 832  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4122469"  
/lab\_lib="NH\_MGC\_17"  
/tissue\_type="rhabdomyosarcoma"  
/lab\_host="PH10B (phage-resistant)"  
/notes="Organ: muscle; Vector: pOTB7; Site: 1; EcoRI;  
Site 2; XhoI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 195 a 256 c 247 g 134 t  
ORIGIN  
Query Match 7.0%; Score 424; DB 11; Length 832;  
Best local Similarity 92.5%; Pred. No. 3.1e-64;  
Matches 557; Conservative 0; Mismatches 30; Indels 15; Gaps 10;  
QY 472 agataacacccgagctcagcttcagaaagggtgtgtttccatggcaacaa 531  
DB 83 AGATAACACCCCGAGTCATCAGCTCTTCGAGAAAGGTGTGTTTCATGCAACAA 142  
QY 532 cagaactaccagcagacttcagaaacatcacccctagagattacagccagct 591  
DB 143 CAGAACTACCAGCAGACTTCGAGAAACATCACCCCTAGAGATTACAGCCGAGT 202  
QY 592 caggccgggttaagtgcagccgagcagcgtctgcgcagagactattatacccgag 651  
DB 203 CAGCGGGGGTAAAGCTGCGACCGGAGCGGCTCTCGCCACAGAGACTATTATACCCGAG 262  
QY 652 ccttaccgagctatgaggtggcgtggcgcgcctctggcactgcagccgctggcc 711  
DB 263 CTTTACCGAGCTATGAGGGTGGCGCTGGCAGCGCCCTCTGGCAGCCGCGTGGCC 322  
QY 712 gccgacaagtaccacgagcagcaagcccttcgcaacacagcagcctcaggggag 771  
DB 323 GCCGACAGTAGTACCACCGGAGCAGCAAGAGCCCTCCGACACAGCAAGCCGTGACGGGAGG 382  
QY 772 cc-ggcttccctggctac-ggcgtccagcagcagccctaccagggcgctagtctg 829

RESULT 14

AZ642061 501 bp DNA GSS 14-DEC-2000  
LOCUS IM0504124R Mouse 10kb plasmid UUGCIM library Mus musculus genomic  
DEFINITION Clone UUGCIM0504124 R. DNA sequence.  
ACCESSION AZ642061  
VERSION AZ642061.1 GI:11768322  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 501)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.,  
and Wright, D., Weiss, R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308 Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: [ddunn@genetics.utah.edu](mailto:ddunn@genetics.utah.edu)  
Insert Length: 10000 Std Error: 0.00  
Plate: 0504 row: 1 column: 24  
Seq primer: CACACAGGAACACGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 501.  
Location/Qualifiers  
1. 501  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGCIM0504124"  
/clone\_lib="Mouse 10kb plasmid UUGCIM library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, Ti-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(<http://www.jax.org/resources/documents/dnares/>). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4

TITLE

JOURNAL

COMMENT

BASE COUNT  
ORIGIN

Query Match	6.5%;	Score 390.6;	DB 13;	Length 501;
Best Local Similarity	86.2%;	Pred. No. 1.9e-56;		
Matches 432; Conservative	0;	Mismatches 69;	Indels 0;	Gaps 0;

QY 1535 ccgaggcgctaacctctctcaaggttccaaacgcgtgcgcgtatgacaaacttgaa 1594  
 |||||  
 Db 1 ctgtagctccctccctcccttcttactgctccaccctctacacactgattgcccaattggagat 60  
 QY 1595 acctccctcaacagcaagacgcgtctcagcaaccggyggtctcccgagaggaactctaac 1653  
 |||||  
 Db 61 actttccctctatagccagcagccgttatgactgagggccttccccaagacatcacagac 120  
 QY 1655 aaagccacttcattgcccctgtctcatccctcccccgaagatgcaacagctctgtgaa 1711  
 |||||  
 Db 121 aaagcactttatgccccctgctttatccctctccccaacgattgtgcacgctctggacac 180  
 |||||  
 QY 1715 ccacagctcggcactcgcgaagcccttcgaagaagaaagcctcccttgagaaactcgtgtccg 1774  
 |||||  
 Db 181 ccagagcccgacactgcagaccctctggaaagaagaaactcctctgacaaacttgcctctccg 240

QY 1775 atccaaacccgcagagagccctcaacggcgctacatctcaagcgtgagagaaactctccaaacccg 1833  
Db 241 acctaaacctcgcagaaactctcaacagctcactacctccaaagctgaaacatctccaaacccg 300  
QY 1835 tccaaacagctgctgctctccaaagctctgctgctgcgcagaaagaaagtgttcaagaacctcg 1894  
Db 301 tgcacacagctcttgctgctccaaacgtacatgtcccaagaagaaagggtgcacaaagacctcg 360  
QY 1895 ttctccagagccccagagagagataaagccagacatgtgagcccccgaagagcgctactc 1955  
Db 361 ttctccagagacctccaaagcagacacaaagccacacactgtgaccttgaaagcagacgacacttt 420  
QY 1955 cagcccgagccccgcagagcacaacgcgtgtctcagagcccgccagagcaacgcgcaagtcacagc 2014  
Db 421 cagctgagaccaaagcgagcacaacgcgctgctgtagagcccgccgaagcagacagccacagtcacacc 480  
QY 2015 atgcgagagccgcagagagagcg 2035  
Db 481 atgctgagaccaaagcagacactcg 501

RESULT	15
AA655431	
LOCUS	613 bp MRNA
DEFINITION	v14c05.r1 stragene mouse heart (#937316) EST 04-NOV-1997
	IMAGE:1111624.5f similar to gb:U29801 Mouse mRNA for unknown cDNA clone
ACCESSION	AA655431 complete cds (MOSBY), mRNA sequence.
VERSION	AA655431.1 GI:2591585
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus

REFERENCE  
AUTHORS

Eukariyota; Metazoa: Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria: Rodentia; Sciurognathi; Muridae; Mus.  
1 (pages 1 to 613)  
Marrs, K., Hillier, J., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Gelsel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,

TITLE The Wasnu-THM Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT Contact: Maria M/Mouse EST Project

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel.: 314.286.1800  
Fax.: 314.286.1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LINT, contact the  
IMAGE Consortium ([info@imgl.nhl.gov](mailto:info@imgl.nhl.gov)) for further information.  
MGI:647968  
Seq primer: -28n3 rev1 ET from Amersham  
High quality sequence stop: 391

FEATURES	Location/Qualifiers
source	1. .613

```

/organism="Mus musculus"
/strain="NIH/Swiss"
/db_xref="taxon:10090"
/clone="IMAGE:1211624"
/clone_id="Stratagene mouse heart (#937316)"
/sex="pooled"
/tissue_type="heart"
/dev_stage="13 day embryos"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: heart; Vector: pBluescript SK-; Site:1;
EcoRI; Site2: XhoI; Cloned unidirectionally. Primer
Oligo dt. 93 pooled NIH/Swiss 13 day embryo hearts.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
adaptor sequence: 5' GAATTCGCGAGG 3' ~3' adaptor
sequence: 5' CTCGAGTTTCTTTTTTTTTT 3' "
```

Query Match	6.2%;	Score 373.4;	DB 10;	Length 613;
Best Local Similarity	81.4%;	Pred. No. 2e-55;		
Matches 456;	Conservative	0;	Mismatches 102;	Indels 2;
			Gaps	2

QY	1716	ccagatgctgcgaactgcgaagccctctcgaagagcaagaatccctgcgaactgcgtctcgga	1777
Dd	5	ccagcccgagactgcgaagccctctgcgaagagcaagaatccctgcgaactgcgtctcgga	64
QY	1776	tctcagccctgcgaagcctcgaagcgctgcgaactaaagttgagaaacattcccaaccgt	1833
Dd	65	cctcagccctgcgaagcctcgaagcgctgcgaactaaagttgagaaacattcccaaccgt	124
QY	1836	ccagaagatgctgctctccgaagctgtgtgcgcgaagaagaagtgctcaagaacctcgt	1899
Dd	125	gcagagctgtctgctctccgaagctgtgtgcgcgaagaagaagtgctcaagaacctcgt	184
QY	1896	gtccagagaccccgagcagcataaagacgcgaactgcagcccgaaagagcgcgtacttc	1955
Dd	185	gtccagagaccccgagcagcataaagacgcgaactgcagcccgaaagagcgcgtacttc	244
QY	1956	agccgagcccgagcagcaccgcgtgttaagagccgcgcgaagcagccacagctccacgca	2011
Dd	245	agccgagcccgagcagcaccgcgtgttaagagccgcgcgaagcagccacagctccacgca	304
QY	2016	tgcgagagccgagagagtcgcgaactactctgagcaggtcccgagagcccaactgagcagagctt	2077
Dd	305	tgccttagccacaggaacacagctacacggaatggctcttagagaccctcttagaccacacttc	364
QY	2076	cctctacatgcgaacagcagcccgctggcagccctgcgaag-gtcaacagcaactcgaagacga	2134
Dd	365	gcctcactcgaagcagcccgctggcagccctgcgaag-gtcaacagcaactcgaagacga	424
QY	2135	agcccgagtcctgtccaactggtctctgtgacactctctcgaagcaatgltcccaacatctgt	2194
Dd	425	agcccgagtcctgtccaactggtctctgtgacactctctcgaagcaatgltcccaacatctgt	484





890 TACCTCAGGGGTGGCCAGTATGAGAGACTTGATGAAAAGACAGCA 939  
 151 ValProProSerArgInTyraGluGlnGlnValProPheAr 167  
 940 GTGCCGCCACGACGAGTATGACAGAGCGGCGCCAGGTGCTTCG 989  
 167 gThrHisSerLeuHisValGlnGlnProProProGlnGlnProLeuA 184  
 990 GATCTACTCCCTCAGCTCAGCAGCCAGCCGCGCGCCAGCAGCCCTGG 1039  
 184 IatProProLysLeuGlnArgGlnLysLeuGlnAsnAspIleAlaSerPro 200  
 1040 CATACCCCAAGCTCCAAAGCGCAGAGCTGCGAAGACACATTGCTCCCT 1089  
 201 LeuProPheProGlnGlnThrHisPheProGlnHisSerGlnSerPhePr 217  
 1090 CTGCCCTTCCGCCAGGGTACCCACTTCTCTCAGCATTCCTCCAGTCCCTCC 1139  
 217 oThrSerSerThrTySerSerValGlnGlnGlyGlyGlnGlyAlaH 234  
 1140 CACTCTCCACCTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1189  
 234 IserTyLysSerCysThrAlaProThrAlaGlnProHisAspArgPro 250  
 1190 ACTCTATAAGAGTTGCACAGCAGCCAGCTGCCAGCCGCCATGACAGCGC 1239  
 251 LeuThrAlaSerSerSerLeuAlaProGlyGlnArgValGlnAsnLeuHi 267  
 1240 CTGACTGCCAGCTCCAGCTTGGCCCGGGGCGAGGGTCCAGATCTCA 1289  
 287 sAlaTyGlnSerGlyArgLeuSerTyAspGlnGlnGlnGlnGlnG 284  
 1290 TGCTTACCAGTGGGGCGGCTCAGTATGACAGCAGCAGCAGCAGCAGC 1339  
 284 InGlnGlnGlnGlnGlnGlnAlaLeuGlnSerArgHisAlaGln 300  
 1340 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1389  
 301 GluThrLeuHisTyGlnAsnLeuAlaLysTyGlnHisTyGlyGlnG 317  
 1390 GAAACCTCTCATACCAAAACCTCCGCCAGTATCAGCATCAGGCGCAGCA 1439  
 317 nGlyGlnGlyTyrcysGlnProAspAlaValArgThrProGlnGlnT 334  
 1440 AGGCCAGAGGCTACTGCCAGCGGCGAGCAGCGCTCCGAGCCGCCAGAGCA 1489  
 334 yTyrcGlnPhePheSerProSerSerHisSerProAlaArgSerVal 350  
 1490 ACTACAGAGCTTCCAGCCCGCAGCTCCAGCCACTCCCGCGCGCTCCGCTG 1539  
 351 GlyArgSerProSerTySerSerThrProSerProLeuMetProAsnLe 367  
 1540 GGCCGCTCAGCTCTTACAGTTCACACCTGCGCGCTGATGCCAAACCT 1589  
 367 uGlnAsnPheProTySerGlnGlnProLeuSerThrGlyAlaPheProA 384  
 1590 GGAGAATCTTCCCTACAGCCAGCAGCGCTCAGCAGCCGCGCGCTTCCCG 1639  
 384 IaglyIleThrAspHisSerHisPheMetProLeuLeuAsnProSerPro 400  
 1640 CAGGATCACTGACACAGCCAGCTTCAATGCCCTCTCAATCCCTCCCA 1689  
 401 ThrAspAlaThrSerSerValAspThrGlnAlaGlyAsnCysLysProLe 417  
 1690 ACGGATGCCAGCTCTGTGGACACCCAGCTGGCAACTGCCAGCCCT 1739  
 417 uGlnLysAspLysLeuProGlnAsnLeuLeuSerAspLeuSerLeuGlnS 434  
 1740 TCAGAGGAGCAAGTCCCTCAGAACTCTCTCTCTCTCTCTCTCTCTCTCTCT 1789  
 434 erLeuThrAlaLeuThrLeuGlnValGlnAsnLysSerAsnThrValGln 450

1790 GCCTCAGCGCGCTGACCTTACAGGTGGAGACATCTCCAAACCGCTCCAG 1839  
 451 GlnLeuLeuLeuSerLysAlaValProGlnLysLysLysValLysAs 467  
 1840 CAGCTGCTGCTCTCCAAAGGTCTGTCGCCAGAGAAAGAGTGTCAAGAA 1889  
 467 nLeuValSerArgThrProGluGlnHisLysSerGlnHisCysSerProG 484  
 1890 CPTGCTGTCAGGACCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1939  
 484 IuGlySerGlyTySerAlaGluProAlaGlyThrProLeuSerGluPro 500  
 1940 AAGGAGCGGCTACTCAGCGAGCCCGGCGCACACCGCTGTACAGCGC 1989  
 501 ProSerSerThrProGlnSerThrHisAlaGluProGlnGlnAlaAspTy 517  
 1990 CCGAGCAGCAGCGCCAGCTTCCAGCATCGCGAGCGCGAGGCGCGACTA 2039  
 517 rLeuSerGlySerGluAspProLeuGluArgSerPheLeuTyrcysAsnG 534  
 2040 CTTGAGCGGCTCCGAGGACCCACTGGAGCGAGCTTCTCTACTGCAACC 2089  
 534 InAlaArgGlySerProAlaArgValAsnSerAsnSerLysAlaLysPro 550  
 2090 AGCCCGGTGGCAGCGCTTCCAGGCTCAACAGCAACTCGAAGGCCAAGCC 2139  
 551 GluSerValSerThrCysSerValThrSerProAspAspMetSerThrLy 567  
 2140 GATCTGCTGCTCCACTGTTCTGTGACCTTCTCTGAGCAGCATGTCCACAA 2189  
 567 sSerAspSerPheGlnSerLeuHisGlySerLeuProLeuAspSerP 584  
 2190 ATCTGACAGCTCTTCCAGAGCTTACAGCGAGCTTCCGCTGCGAGCT 2239  
 584 heSerLysPheValAlaGlyGluArgAspCysProArgLeuLeuSer 600  
 2240 TCTCAAAGTCTGTCGGGTGAGCGGAGCTTCCGCGGCTGCTGCTCAGC 2289  
 601 AlaLeuAlaGlnGluAspLeuAlaSerGluIleLeuGlyLeuGlnGluAl 617  
 2290 GCCTTGGCAGGAGGAGCTTGGCTCCGAGATCTTGGGCTGCGAGGAGC 2339  
 617 alledGlyGluLysAlaAspLysAlaTrpAlaGluAlaProSerLeuVal 634  
 2340 CATCGGTGAGAAGCGCCAGAAAGCTTGGGCTGAGCAGCAGCAGCTGTGCA 2389  
 634 yAspSerSerLysProProPheSerLeuGluAsnHisSerAlaCysLeu 650  
 2390 AGCAGCAGCAGCAAGCCACCTTCTGCTGGAGAACCCAGCGCTGCTGCTG 2439  
 654 AspSerValAlaLysSerAlaTrpProArgProGlyGluProGluAlaLe 667  
 2440 GACTCTGTCGCAAGAGTGGTGGCCCGGCTGGGAGCCCGAGGCGCT 2489  
 667 uProAspSerLeuGlnLeuAspLysGlyGlyAsnAlaLysAspPheSerP 684  
 2490 GCCCGACTCTTGGCAGCTGGCAGAGCGGCAATGCCAAGGACTTCAGCC 2539  
 684 roGlyLeuPheGluAspProSerValAlaPheAlaThrProAspProLys 700  
 2540 CAGGGCTGTTGAGACCTTCTGCTGGCTTCTGCTGCTGCTGCTGCTGCTGCT 2589  
 701 LysThrThrGlyProLeuSerPheGlyThrLysProThrLeuGlyValPr 717  
 2590 AAGACAATGCT 2639  
 717 oAlaProAspProThrThrAlaAlaPheAspCysPheProAspThrThra 734  
 2640 TGCTCCAGACCCCTACAGCAGCTTTGACTGTTTCCCGGACACACCG 2689  
 734 laAlaSerSerAlaAspSerAlaAsnProPheAlaTrpProGluGluAsn 750  
 2690 CTGCCAGCTCAGCGGAGCAGCGCCACACCTTTCCTGCTGGCGAGGAAAC 2739

751 LeuGlyAspAlaCysProArgTrpGlyLeuHisProGlyGluLeuThrIly 767  
767 sglLeuGluGluGlnGlyGlyLysAlaSerAspGlyLleSerLysGlyAspTr 784  
2740 CTGGGGGATGCTTGTCCAGAGTGGGATTCACCTGGCGAGACTTACAA 2789  
767 sglLeuGluGluGlnGlyGlyLysAlaSerAspGlyLleSerLysGlyAspTr 784  
2790 GGGCTGAGACAGAGTGGAGAGGCTCCAGATGATCAGTCAAGAGGAGCA 2839  
784 hrHisGluAlaSerAlaCysLeuGlyPheGlnGlnGluAspProGly 800  
2840 CCATGAGGCTTCGGCTGCTGGCTTCAGAGAGAGAGCCCTGGG 2889  
801 GluLysValAlaSerLeuProGlyAspPheLysGlnGlnGluValGly 817  
2890 GAGAGAGTGGCTGCTGGCTTCGGGAGCTTCAAGCAGAGAGAGTGGTGG 2939  
817 YValLysGlnGluAlaGlyLysLeuGlnGlyCysProGluValAlaLys 834  
2940 GGTAGAGAGAGAGAGTGGCTGCTGCTGAGTCCCGAGTGGCGCAAG 2989  
834 laAspArgTrpLeuGluAspSerArgHisCysCysSerThrAlaAspPhe 850  
2990 CTGACCGGTGGCTGAGAGACAGCCGGCAGCTGCTTCCAGCCGACCTTC 3039  
851 GlyAspLeuProLeuLeuProProThrSerArgPheGluAspLeuAla 867  
3040 GGGAGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3089  
867 agLugLugLugLysSerSerLeuGlyLeuLeuGlySerProGluGln 884  
3090 TGAAGAGAGAGTACTCTCTCTATGAGCTCTGAGCAGCCCGAGCAGA 3139  
884 rgrProGlyMetGluAspProLeuSerProLysAlaProLeuLleCysThr 900  
3140 GGGCTGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3189  
901 LysGlnGluValGlnGluValAlaLeuAspSerLysAlaGlyTrpGlySer 917  
3190 AAGAGAGAGTGGAGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 3239  
917 ccYHisLeuSerGlyLysSerValLleLeuLeuGlyProThrValGly 934  
3240 GTGGACCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3289  
934 hrGluSerLysValGlnSerTrpPheGluSerSerLeuSerHisMetLys 950  
3290 CCGAGTCAAGGTCCAGAGAGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3339  
951 ProGlyGlnGluGlyProAspGlyGlnArgAlaProGlyAspSerThr 967  
3340 CCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3389  
967 rSerAspAlaSerLeuAlaGlnLysProAsnLysProAlaValProGlu 984  
3390 CTGGAGAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3439  
984 laProLleAlaLysLysGluProValProGlyLysSerLeuArgSer 1000  
3440 CGCCATCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3489  
1001 ArgArgValHisArgGlyLeuProGluAlaGlnAspSerProCysArg 1017  
3490 GGTGGGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3539  
1017 aProValLeuProLysAspLeuLeuProGlyLysCysThrGlyProp 1034  
3540 ACCAGTGTGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3589  
1034 rGlnGlnGlnMetGluValGlnGlyAlaProGlyLysArgValAspSer 1050  
3590 CCAAGGAG 3639

1051 GlyLeuProArgMetCysThrArgSerLeuThrAlaLeuSerGluPro 1067  
3640 GGGCTCCAG 3689  
1067 gThrProGlyProProGlyLeuThrThrProAlaProAspLysAl 1084  
3690 CAGGCGGAG 3739  
1084 euGlyGlyLysGlnArgAlaAlaPheLysSerGlyLysArgValGly 1100  
3740 TGGGGGAG 3789  
1101 ProSerProLysAlaAlaSerSerProSerAspProAlaAlaLeuPro 1117  
3790 CCTTCCAG 3839  
1117 laAlaSerAspSerProMetGlySerLysThrLysGlnThrAspSer 1134  
3840 GGGCTCCAG 3889  
1134 rSerThrProGlyLysAspGlnArgSerMetLleLeuArgSerArgThr 1150  
3890 CCAAG 3939  
1151 LysThrGlnGlnLlePheHisSerLysArgArgProSerGluGlyArg 1167  
3940 AAAAG 3989  
1167 gLeuProAsnCysArgAlaThrLysLysLeuLeuAspAsnSerHisLeu 1184  
3990 GCTCCAG 4039  
1184 roAlaThrPheLysValSerSerSerProGlnLysGlnLysValSer 1200  
4040 CCGCCAG 4089  
1201 GlnArgAlaArgValProLysProGlyValAlaGlySerLysLeuSerAsp 1217  
4090 CAGGAG 4139  
1217 gProLeuHisAlaLeuLysArgLysSerAlaPheMetAlaProValPro 1234  
4140 GCGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4189  
1234 hrLysLysArgAsnLeuValLeuArgHisGlySerSerSerSerAsn 1250  
4190 CCAAG 4239  
1251 AlaSerAlaMetGlyLysMetGlyArgArgArgGlyLeuArgValPro 1267  
4240 GCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4289  
1267 oProSerSerArgLysCysLeuLeuProArgLysProSerProProArg 1284  
4290 ACCCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4339  
1284 laMetAlaSerLeuProGlnSerSerHisProArgArgProProPheLeu 1300  
4340 CAATGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4389  
1301 ProGlnAlaArgLeuSerAlaAlaPheGlnGlyAlaMetLysThrLysVal 1317  
4390 CTTCAAG 4439  
1317 lleuProArgLysGlyArgGlyLeuLysLeuGlnAlaLleValGln 1334  
4440 GCTGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4489  
1334 ysLleThrSerProSerLeuLysPheAlaCysLysAlaProGlyVal 1350  
4490 AGATCACTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4539  
1351 SerProGlyAsnProLeuSerProSerLeuSerAspLysAspArgGlyLe 1367

```
|||||
4540 TCTCTCGGTAATCTCTGAGCCCATCCCTTTCCGACAAAGACCGTGGGCT 4589
|||||
1367 uLysGlyAlaGlySerProValGlyValGluGluGlyLeuValAsnV 1384
|||||
4590 CAGGGTGCTGGGGCGACGCCACTGGGGGTGGAGAGGGCTGGTAAATG 4639
|||||
1384 alGlyThrGlyGlnLysLeuProThrSerGlyAlaAspProLeuCysArg 1400
|||||
4640 TGGGACCGGGGAGAGGCTCCCAACTCTGGGGCTGATCCGTTATGCAGA 4689
|||||
1401 AsnProThrAsnArgSerLeuLysGlyLysLeuMetAsnSerLysLysLe 1417
|||||
4690 AATCCACCAACAGATCTTAAAGGCAACATCATGACAGTAAAGAACT 4739
|||||
1417 userSerThrAspCysPheLysThrGluAlaPheThrSerProGluAlaL 1434
|||||
4740 GTCCTCTACTGACTCTTTAAACCGAGGCTTCACATCCCGGAGGCC 4789
|||||
1434 euGlnProGlyThrAlaLeuAlaProLysLysArgSerArgLysGly 1450
|||||
4790 TCAGGCTGGGGGAGCTGCCCTGGCGCTAAGAGAGGAGCGGGAAGGC 4839
|||||
1451 ArgAlaGlyAlaHisGlyLeuSerLysGlyProLeuGlnLysArgProTy 1467
|||||
4840 CGGGCAGGGGCCCTGAGCTCTCCAAAGGCCCGCTGGAGAGCGGCCCTA 4889
|||||
1467 rLeuGlyProAlaLeuLeuThrProArgAspArgAlaSerGlyThrG 1484
|||||
4890 TCTTGGCCGGCTCTCTGATCTCCCGAGACAGGGCCAGTGGCACAC 4939
|||||
1484 lngAlaSerGluAspAsnSerGlyGlyGlyLysLysProLysMet 1500
|||||
4940 AAGGGCCAGTGGAGCAACTCTGGTGGAGGAGGCAAGAGCAAGATG 4989
|||||
1501 GluGluLeuGlyProAlaSerGlnProProGluGlyArgProCysGlnPr 1517
|||||
4990 GAGGAGCTGGGCCCTGCTCCAGCCCGCCGAGGGGAGGCGCTGCCAGCC 5039
|||||
1517 oGlnThrArgAlaGlnLysGlnProGlyHisThrAsnTySerSerTyrs 1534
|||||
5040 CCAGCAAGGCGCAGAAACAGCCAGGCGCCACCACTACAGCACTATT 5089
|||||
1534 erLysArgLysArgLeuThrArgGlyArgAlaLysAsnThrThrSerSer 1550
|||||
5090 CCAAGCGGAAGCGCTCACTCGGGGCGGGCCGAGAACACCACTCTTCA 5139
|||||
1551 ProCysLysGlyArgAlaLysArgArgGlnGlnGlnValLeuProLe 1567
|||||
5140 CCTGTAGGGGCGTGCAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5189
|||||
1567 uAspProAlaGluProGluLeuArgLeuLysTyrrLysSerCysLysA 1584
|||||
5190 GGATCCCGCAGAGCTGAAATCCGCTCAAGTACATTTCTCTTCCACAGC 5239
|||||
1584 rGluArgSerAspSerArgThrProAlaPheSerProPheValArgVal 1600
|||||
5240 GCTGTAGGTCACAGAGCGGAGCCCGCTCTCCCTCTCCCTCTGGGGGTG 5289
|||||
1601 GluLysArgAspAlaPheThrThrLecCysThrValValAsnSerProG 1617
|||||
5290 GAGAGCGAGAGCGCTTCCACCATATGCATGTGTCAACTCCCTCGG 5339
|||||
1617 yAspAlaProLysProHisArgLysProSerSerSerAlaSerSerSerS 1634
|||||
5340 AGATGCGCCCAAGGCCCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5389
|||||
1634 erSerSerSerSerPheSerLeuAspAlaAlaGlyAlaSerLeuAlaThr 1650
|||||
5390 CATCTGCTGCTCTCTCTCTCTGAGGAGGAGGAGGAGGAGGAGGAGG 5439
|||||
1651 LeuProGlyGlySerIleLeuGlnProArgProSerLeuProLeuSerSe 1667
|||||
```

```
5440 CTCCCTGAGGCTCCATCCTGCAGCCGCGGCCCTCCTTGCCTCTCTCCTC 5489
1667 rHmMerHisLeuGlyProValValSerLysAlaLeuSerThrSerCysL 1684
5490 CACGATGCACTTGGGGCTGTGGTTTCCAGAGCCCTGAGTACCTCTTGCC 5539
1684 euValCysCysLeuCysGlnAsnProAlaAsnPheLysAspLeuGlyAsp 1700
5540 TTGTTTGTGCTCCTCTGCCAAACCGGCGCAACTTCAAGGACCTTGGGG 5589
1701 LeuCysGlyProTyrrProGluHisCysLeuProLysLysLysProLy 1717
5590 CFTCTGGGGCTACTACCTGACACTGCTCCCTCCCAAAAGAGAGCCCAA 5639
1717 sLeuLysLysValArgProGluGlyThrCysGluGluAlaSerLeuP 1734
5640 ACTCAGAGAGAGGTGGCGGCAAGAGCACCTGTGAGGAGGCGCTCGCTG 5689
1734 roLeuGluArgThrLeuLysGlyProGluCysAlaAlaAlaThrAla 1750
5690 CGTTGAGAGAACACTCAAGGTCCGAGTGTGAGTGCAGCTGCCGCCACTGCC 5739
1751 GlyLysProProArg 1765
5740 GGAAGCGCCCGCCAGG 5754
seq_name: gb_htg:HS253P07
seq_documentation_block:
LOCUS HS253P07 160544 bp DNA HTG 20-APR-2001
DEFINITION Homo sapiens chromosome 17 clone PAC RPI-1 253P07 map 17p11.2, ***
SEQUENCING IN PROGRESS ***, 2 ordered pieces.
ACCESSION AL354000.6 GI:13752103
VERSION AL354000.6 GI:13752103
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 160544)
AUTHORS Ramser,J., Langer,I., Steffens,C., Klein,M., Borzym,K., Junker,E.,
Marquardt,I., Moll,K., Radelof,U., Francis,F., Seranski,P.,
Poustka,A., Reinhardt,R. and Lehrach,H.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 160544)
AUTHORS MOLGENR.
TITLE Direct Submission
JOURNAL Submitted (28-APR-2000) MOLGENR, Abt. Lehrach, Max Planck Institut
Fuer Molekulare Genetik, Ihnestrasse 73, Berlin, 14195 Germany
On May 14, 2001 this sequence version replaced gi:9931105
gi:10280527.
contig 01 1. 86093
contig 02 86194. 160544
Clone received from the Resource Centre of the Human Genome Project
at the Max-Planck-Institute for Molecular Genetics.
* NOTE: this is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 86093: contig of 86093 bp in length
* 86094 86193: gap of 100 bp
* 86194 160544: contig of 74351 bp in length.
Location/Qualifiers
1. 160544
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17p11.2"
FEATURES
source
```



```
/clone="PAC RPCI-1 281113"  
/clone_lib="RPCI1,3-5 Human PAC library, originating  
institute: Roswell Park Cancer Institute, creator: Pieter  
de Jong, P. Ioannou"  
/note="region between markers D17S842-D17S953"  
1  
misc_feature  
/note="T7_end:PAC RPCI-1 253P07"  
160544  
misc_feature  
/note="SP6_end:PAC RPCI-1 281113"  
BASE COUNT 35434 a 44179 c 44032 g 36798 t 101 others  
ORIGIN  
  
alignment_scores:  
Quality: 9191.00 Length: 1758  
Ratio: 5.249 Gaps: 3  
Percent Similarity: 99.602 Percent Identity: 99.545  
  
alignment_block:  
US-09-508-821B-6 x HS253P07/rev  
  
Align seg 1/1 to reverse of: HS253P07 from: 1 to: 160544  
  
1 MetGlnSerPheArgGluArgCysGlyPheHisGlyLysGlnGlnAsnTyr 17  
109850 ATGCAAGCTTTTCGAGAAAGGTGGTTTCATGGCAACACACAGACTA 109801  
17 GlnGlnThrSerGlnGlnThrSerArgLeuGlnAsnTyrArgGlnPro 34  
109800 CCAGCAGACCTCGCAGAGAAACATCACCCCTAGAGAAATACAGGAGCCGA 109751  
34 ArgGlnAlaGlyLeuSerCysAspArgGlnArgLeuLeuAlaLysAspTyr 50  
109750 GTCAGAGCCGGGGCTAAGCTGACGACCGGAGCGGCTGCTGCCAAGACTAT 109701  
51 TyrAsnProGlnProTyrProSerTyrGlnGlyAlaGlyTyrProSer 67  
109700 TATAACCCGAGGCTTACCCGAGCTATGAGGGTGGGGCTGGCAGCCCTTC 109651  
67 GlnTyrAlaAlaAlaValAlaAlaAspLysTyrHisArgLysLysLys 84  
109650 TGGCACTGCAAGCCGGTGGCCCGGACAGTACACCGAGGAGCAGCAAG 109601  
84 LeuProThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 100  
109600 CCTGCCACACAGCAAGGCTGAGGGGAGCGCGCTTCCCTGGGCTAC 109551  
101 GlyValAlaGlnAspSerSerProTyrProGlyArgTyrAlaGlnGln 117  
109550 GCGGTCCAGGACAGACAGCCCTTACCCAGGCGGCTATGCTGGTGAAGAG 109501  
117 GlnGlnAlaThrProGlyAlaProGlnProProProProGlnProGlnPro 134  
109500 CCTTGAAGCTTGGGGGCCCCCAGACCAACCCCAAGCCGAGCCGAC 109451  
134 LeuProAlaGlyValAlaLysTyrAspLysLysLysLysLysLysLys 150  
109450 TACCTGCAAGGGGTGGCCAGTATGATGAGAACTTGATGAAAAGACAGCA 109401  
151 ValProProSerArgGlnTyrAlaGlnGlnGlnGlnGlnGlnGlnGln 167  
109400 GTGCCCCCAGCAGGAGCAGTATGACAGGAGGAGGAGGAGGAGGAGGAG 109351  
167 GlnHisSerLeuHisValGlnGlnProProProProProGlnGlnProLeu 184  
109350 GACTCACTCCCTGCACTCCAGACAGCCAGCCGCGCCAGCAGCCCTGG 109301  
184 LeuTyrProLysLeuGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGln 200  
109300 CATACCCCAAGGCTCCAAAGCAGAAAGCTGAGACGACATCTCCCTCC 109251  
201 LeuProPheProGlnGlnGlnTyrHisPheProGlnHisSerGlnSerPhe 217
```

108350 CCGAGCAGCAGCCACAGTCCAGCATCGGAGCCCGAGGAGCCGACTA 108301  
517 rLeuSerGlySerGluuaspProLeuGluArgSerPheLeuTyrCysAsnG 534  
108300 CCGAGCAGCCGCTCGAGGACCCACTGAGCGCAGCTCTCTACTGCAAC 108251  
534 lnaLaArgGlySerProAlaAaGValAsnSerAsnSerLysAlaLysPro 550  
108250 AGGCGCGTGGCAGCCCTGCGAGGTCACAGGTCACAGCAACTCGAGGCCAAGCC 108201  
551 GluSerValSerThrCysSerValnSerProAspMetSerThrly 567  
108200 GAGTCGCTGTCACCTCTCTGACTCTCTGACCTCTCTGACGACATGTCACCAA 108151  
567 sSerAspSerPheGluSerLeuHisGlySerLeuProLeuAspSerP 584  
108150 ATCTGACGACTCTTCAGAGCTACAGGAGCTGCGGCTGACAGCT 108101  
584 heSerLysPheValAlaGlyArgAspCysProArgLeuLeuSer 600  
108100 TCTCCAAGTTCGTGGCGGTGAGCGGACTGTCCGCGGCTGCTCAGC 108051  
601 AlaLeuAlaGlnLeuaspLeuAlaSerGluLeuGlyLeuGlnGluAl 617  
108050 GCCCTGGCAGAGGAGCTTGGCTCCAGATCTCTGGGCTGCGAGGAGC 108001  
617 alleGlyGluLysAlaAspLysAlaTrpAlaGluAlaProSerLeuVal 634  
108000 CATCGGTGAGAGGCCACAAAGCTTGGGCTGAAGCACCCAGCTGGTCA 107951  
634 ysAspSerSerLysProPheSerLeuGluAsnHisSerAlaCysLeu 650  
107950 AGGACAGCAGCAAGCCACCTCTCGCTGGAGAACACAGCGCTGCTG 107901  
651 AspSerValAlaLysSerAlaTrpProArgProGlyGluProGluAlaLe 667  
107900 GACTCTGTGCCAGAGCTGGTGGCCCGGCTGGGAGCGGAGCCCT 107851  
667 uproAspSerLeuGlnLeuAspLysGlyGlyAsnAlaLysAspPheSerP 684  
107850 GCCCGACTCTCTGACAGTGGACAGGCGGCAATGCCAAGGACTTCAGCC 107801  
684 roGlyLeuPheLeuaspProSerValAlaPheAlaThrProAspProLys 700  
107800 CAGGCTGTTCAGAGCCCTTCGGTGGCTTCGCTACGCTGACCCCAA 107751  
701 LysThrThrGlyProLeuSerPheGlyThrLysProThrLeuGlyValPr 717  
107750 AAGCAACTGTGCTCTCTCTTGGTACCAAGCCACCCCTTGGGTTCC 107701  
717 oAlaProAspProThrThrAlaAlaPheAspCysPheProAspThrThra 734  
107700 TGCTCCAGACCCCACTACAGCAGCTTTTGACTGTTTCCCGGACACACCG 107651  
734 laAlaSerSerAlaAspSerAlaAsnProPheAlaTrpProGluAsn 750  
107650 CTGCCAGCTCAGCGGACAGCGCCACCCCTTTCCTGCGCCAGAGGAAAC 107601  
751 LeuGlyAspAlaCysProArgTrpGlyLeuHisProGlyGluLeuThrly 767  
107600 CTGGGGGATGCTTTCAGGTGGGATTCGACCCCTGGGAGCTTAOCAA 107551  
767 scLyLeuGlnGlnGlyGlyLysAlaSerAspGlyLeuSerLysGlyAspT 784  
107550 GGGCTGGAGCAGGCTGGAGGCTTCAGATGCGATGCATCACAAGGGGACA 107501  
784 hrHisGluAlaSerAlaCysLeuGlyPheGlnGluLeuaspProProGly 800  
107500 CCAATGAGCTTCGGCTGCTGCTGCTTCCAGAGGAGGAGCCGCCCTGG 107451  
801 GluLysValAlaSerLeuProGlyAspPheLysGlnGluValGly 817  
107450 GAGAGGCTGGCTCTGTTGCCCGGGGACTTCAGCGAGGAGGAGTGGTGG 107401

817 yValLysGlnGluAlaGlyLeuLeuGlnCysProGluValAlaLysA 834  
107400 GGTGAAGAGAGAGCCAGTGGCTGCTGAGTGCCTCCGAGGTGGCCAAAG 107351  
834 laAspArgTrpLeuGluaspSerArgHisCysSerThrAlaAspPhe 850  
107350 CTGACCGGTGGCTGGAGGACAGCGGCTGCTGTTCCACGCGGACTTC 107301  
850 GlyAspLeuProLeuLeuProThrSerArgLysGluaspLeuGluAl 867  
107300 GGGGACCTTCCACTGCTGCACCCACAGCAGAAAGAGGAGCTGGAAGC 107251  
867 aGluGluGlyTrpSerSerLeuCysGluLeuLeuGlySerProGluGlnA 884  
107250 TGAGAGGAGTACTCTCTCTATGTGAGCTTCTGGGAGCCGCGGAGCAGA 107201  
884 rGProGlyMetGlnAspProLeuSerProLysAlaProLeuLeuLeuCysThr 900  
107200 GGCCTGGCATGAGGACCGCTGTACCCAAAGCCCACTCATCTGCACC 107151  
900 LysGluGluValGluGluValLeuaspSerLysAlaGlyTrpGlySerPr 917  
107150 AAGGAGGAGTGGAGGAGTGTGACTCCAAAGCCGCTGGGCTGCTCC 107101  
917 oCysHisLeuSerGlyGluSerValLeuLeuLeuGlyProThrValGlyT 934  
107100 GTGCCACTCTCAGGAGTGTGCTCATCTGCTGGGCTTACAGTGGCA 107051  
934 hrGluSerLysValGlnSerTrpPheGluSerSerLeuSerHisMetLys 950  
107050 CGGAGTCAAGGCTCAGAGTGGTGTGAGTCTCTCTCATACATGAAG 107001  
951 ProGlyGluGluGlyProAspGlyGluArgAlaProGlyAspSerThrTh 967  
107000 CCAGGTGAGAGGGGCTGTGTGGGAGCGAGCTCCAGGGGATTCACAC 106951  
967 rSerAspAlaSerLeuAlaGlnLysProAsnLysProAlaValProGluA 984  
106950 CTCGGAGCCTCTCTGCGCAGAAAGCCCTGCGCAGGCAAGAGCTGCTG 106901  
984 laProLeuAlaLysLysGluProValProArgGlyLysSerLeuArgSer 1000  
106900 CGCCCATCCCAAGAAAGAGCTGTGCGCAGGCGCAAGAGCTTACGGAGC 106851  
1000 ArgArgValHisArgGlyLeuProGluAlaGluaspSerProCysArgAl 1017  
106850 CGTGGGTGCAAGCGGCTGCGCAGGCGAGGACTCCCATGAGGAGC 106801  
1017 aproValLeuProLysAspLeuLeuLeuProGluSerCysThrGlyProp 1034  
106800 ACCAGTGTGCCAAAGAGCTTGTCTCCCTGAATCCTGCACAGGCGCC 106751  
1034 roGlnGlyGlnMetGluGlyValAlaGlyAlaProGlyArgGlyAlaSerGlu 1050  
106750 OCCAGGACAGATGAAGGGCTGGAGCCCGAGCGGGGGCTCGGA 106701  
1051 GlyLeuProArgMetCysThrArgSerLeuThrAlaLeuSerGluProAr 1067  
106700 GGCTCCCGCAGATGTACTGCTCTCTCAGCGGCTGAGTGGAGCCCG 106651  
1067 gThrProGlyProGlyLeuThrThrProAlaProAspLysL 1084  
106650 CAGCGCGGAGCCCGAGGCTGACCCACCCCTGCACCCCGCAGCAAAAC 106601  
1084 euGlyLysGlnArgAlaAlaPheLysSerGlyLysArgValGlyLys 1100  
106600 TGGGGGCAAGCAGCAGCGGCTTCAAGTCGGCAAGCGGTGGGAAG 106551  
1101 ProSerProLysAlaAlaSerSerProSerAsnProAlaAlaLeuProVa 1117  
106550 CCTCACCAGGCTGCTTCCAGCGCCCGCAGCAACCGCGCGCTGCTCTGT 106501

1117 AlaSerAspSerSerProMetGlySerLysThrLysGluThrAspSer 1134  
106500 GGCCTCCGACAGCAGCCCGATGGGCTCCAGACCAAGGAGAGACCTCAC 106451  
1134 roserThrProGlyLysAspGlnArgSerMetIleLeuArgSerThr 1150  
106450 CCAACAGCCCTGGCAAGACCAAGCTGCTCAATGATCTCTGGTCAAGCACC 106401  
1151 LysThrGlnIlePheHisSerLysArgArgArgProSerGluGlyArg 1167  
106400 AAAACCCAGAGATCTTCCACTCCAGAGGGGAGGCGCTCTGAGAGGCGG 106351  
1167 glenProAsnGlyArgAlaThrLysLysLeuLeuAspAsnSerHisLeuP 1184  
106350 GCTCCCAACTGCTGCTCCACAGAGAGCTCTCTGACACAGCAGCCTTGC 106301  
1184 rolaThrPheLysValSerSerSerProGlnLysGluGlyArgValSer 1200  
106300 CCGCCACATTTCAAGGCTCCAGAGAGCCCGAAGAGGAGGAGGCTGAGC 106251  
1201 GlnArgAlaArgValProLysProGlyAlaGlySerLysLeuSerAspArg 1217  
106250 CAGGGGCAAGGGTCCCAAACTGCTGAGGAGGAGCAGCAGCTCTGAGC 106201  
1217 gProLeuHisAlaLeuLysArgLysSerAlaPheMetAlaProValPro 1234  
106200 GCCCTCCATGGCTCAAAAGAGAGTGGGCTTCATGGGGCGGCTCCCA 106151  
1234 hrLysLysArgAsnLeuValLeuAlaGlyHisGlySerSerSerSerAsn 1250  
106150 CCAAGAGCGGAGACCTGCTTCCGAGAGCCGACAGCAGCAGCAGCAGC 106101  
1251 AlaSerAlaMetGlyGluMetGlyArgArgArgGlyLeuArgValProp 1267  
106100 GCCAGTGGCAATGGGGGAGATGGGAGAGGAGGAGGCTGAGAGGTCC 106051  
1267 roProSerSerArgGlyCysLeuLeuProArgLysProSerProProArg 1283  
106050 CACCTCTTCAAGAGATGCTCTCCCAAGAGAGGAGGAGGAGGAGG 106001  
1284 AlaMetAlaSerLeuProGlnSerSerHisProArgArgProProPheL 1300  
106000 GCATGGGAGGCTGGCCAGCAAACTCCACCCCGGAGAGCCCGCCATGCC 105951  
1300 euProGlnAlaArgLeuSerAlaAlaPheGlnGlyAlaMetLysThrLys 1316  
105950 TGCTCAAGCTGCTCTCGGAGGCTTCCAGGGGAGGAGGAGGAGGAG 105901  
1316 sValLeuProProArgLysGlyArgGlyLysLeuLysGluAlaIleValG 1333  
105900 GGTCTGCCAGCCCGGAGAGGCGGAGGCTGAGAGCTGAGAGCATGCTGC 105851  
1333 LysLysIleThrSerProSerLeuLysLysPheAlaCysLysAlaProGly 1349  
105850 AGAAGATCACTCGCCAGGCTCAGAGAGTTCGATGTAAGCGGAGGAG 105801  
1350 AlaSerProGlyAsnProLeuSerProSerLeuSerAspLysAspArgLys 1366  
105800 GCCTCTCTGATATCTCTGAGCCATCTCTTCCAGCAAGAGAGGAGG 105751  
1366 LysLeuLysGlyAlaGlySerProValGlyAlaGluGluGlyLeuValA 1383  
105750 GCTCAAGGCTGCTGGGGGAGCCAGTGGGGGTGGAAGAGGCTGGTAA 105701  
1383 snValGlyThrGlyLysLysLeuProThrSerGlyAlaAspProLeuLys 1399  
105700 ATGTGGGACCGGGGAGAGCTCCCAACTCTGGGGGTATCCGTTAAGC 105651  
1400 ArgAsnProThrAsnArgSerLeuLysGlyLysLeuMetAsnSerLysLys 1416  
105650 AGAATCCACCAACAGATCTTAAAGGCAACTGATGAGAGAGTAAAGAA 105601  
1416 sleuSerSerThrAspCysPheLysThrGluAlaPheThrSerProGluA 1433  
105551 ACTGCTTCTACTGACTGTTTCAAAACCGAGGCTTCAACATCCCGGAGG 105501  
1433 IalaLeuGlnProGlyLysThrAlaLeuAlaProLysLysArgSerArgLys 1449  
105500 CCTGACAGCTGGGGGAGTCCCTGGGAGGCTTCAAGAGGAGGAGGAGG 105451  
1450 GlyArgAlaGlyAlaHisGlyLeuSerLysGlyProLeuGluLysArgPr 1466  
105450 GCGGAGGAGGAGGAGGAGGAGTCCCAAGAGGAGGAGGAGGAGGAGG 105401  
1466 cTyLeuGlyProAlaLeuLeuThrProArgAspArgAlaSerLysP 1483  
105400 CTATCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 105351  
1483 hrGlnGlyAlaSerGluAspAsnSerGlyGlyGlyLysLysProLys 1499  
105350 CACAAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 105301  
1500 MetGluGluLeuGlyProAlaSerGlnProProGluGlyArgProCysG 1516  
105300 ATGAGAGAGCTGGGCTGGGCTCCAGAGGAGGAGGAGGAGGAGGAG 105251  
1516 nProGlnThrArgAlaGlnLysGlnProGlyHisThrAsnLysSerSer 1533  
105250 GCCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 105201  
1533 YrSerLysArgLysArgLeuThrArgGlyArgAlaLysAsnThrThrSer 1549  
105200 ATTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 105151  
1550 SerProCysLysGlyArgAlaLysArgArgGlnGlnGlnValLeuP 1566  
105150 TCACCTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 105101  
1566 cLeuAspProAlaGlnProGlnIleArgLeuLysLysLysSerCysL 1583  
105100 CCGGATGCCGAGAGGCTGAGAAATCCGCTCAAGTCAATTCCTTGGCA 105051  
1583 YsArgLeuArgSerAspSerArgThrProAlaPheSerProPheValArg 1599  
105050 AGCGGCTGAGGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 105001  
1600 ValGluYsArgAspAlaPheThrThrIleCysThrValAlaAsnSerPr 1616  
105000 GTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 104951  
1616 cGlyAspAlaProLysProHisArgLysProSerSerSerAlaSerSer 1633  
104950 TGGAGATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 104901  
1633 eSerSerSerSerSerPheSerLeuAspAlaAlaGlyAlaSerLeuAla 1649  
104900 CCGATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 104851  
1650 ThrLeuProGlyLysSerLysLeuGlnProArgProSerLeuProLeuSe 1666  
104850 ACACCTCCGAGAGGCTCACTGAGAGGAGGAGGAGGAGGAGGAGGAG 104801  
1666 rSerThrMetHisLeuGlyProValSerLysAlaLeuSerThrSerC 1683  
104800 CTCACAGATGCACTGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 104751  
1683 YsLeuValCysCysLeuCysGlnAsnProAlaAsnPheLysAspLeuGly 1699  
104750 GCCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 104701  
1700 AspLeuCysGlyProLysThrProGlnHisCysLeuProLysLysPr 1716  
104700 GACCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 104651  
1716 cLysLeuLysGluLysValArgProGluGlyThrCysGluGluAlaSerL 1733

104700 ABACTCAAGGAGAGGTGGCGGACGAGCAAGCACTGTGAGGAGCCTGCG 104651

1733 eurl-leucluarthThrLeuLyGlyProGluCysAlaAlaAlaAlaThr 1749

104650 TGGCGTGTAGAGAACACTCAAGAGTCCGAGTGTGAGCTGCGCCGCACT 104601

1750 AlagLyLysProArg 1755

104600 GCGGGAAGCCCCCAGG 104583

seq\_name: gb.pr: HSA271791

seq documentation\_block:  
LOCUS HSA271791 23843 bp DNA PRI 17-JUN-2001  
DEFINITION Homo sapiens partial RAIL gene for retinoid-acid induced protein 1,  
exons 1-7.  
ACCESSION AJ271791  
VERSION AJ271791.1 GI:12053794  
KEYWORDS RAIL gene; retinoid-acid induced protein 1.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 23843)  
AUTHORS Seranski, P., Hoff, C., Radelof, U., Hennig, S., Reinhardt, R.,  
Schwartz, C.E., Heiss, N.S. and Poustka, A.  
TITLE RAIL is a novel polyglutamine encoding gene that is deleted in  
Smith-Magenis syndrome patients  
JOURNAL Gene 270 (1-2), 69-76 (2001)  
MEDLINE 21297181  
REFERENCE 2 (bases 1 to 23843)  
AUTHORS Seranski, P.  
JOURNAL Direct Submission  
TITLE Submitted (31-JAN-2000) Seranski P., Molecular Genome Analysis,  
Deutsches Krebsforschungszentrum Heidelberg, Im Neuenheimer Feld  
280, 69120 Heidelberg, GERMANY

FEATURES  
source Location/Qualifiers

1. 23843  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/map="17p11.2"  
join(9590..10205,10271..10806,10881..13830,14024..14992,  
15049..15152,18679..18809,20395..>20488)  
/gene="RAIL"  
9590..20488  
/gene="RAIL"  
9590..10205  
/gene="RAIL"  
/number=1  
join(9590..10205,10271..10806,10881..13830,14024..14992,  
15049..15152,18679..18809,20395..>20488)  
/gene="RAIL"  
/codon\_start=1  
/product="retinoid-acid induced protein 1"  
/protein\_id="CAC20424.1"  
/db\_xref="GI:12053795"  
/translation="MQSFRCGPHGKQYQOCTSOETSLENVROPQAGLSDCROR  
LIAKDYINPOYSYEGGAGTSGTAARAADKIHGSKALPTQGLGHPAPGCVG  
QSSSTPGRYAGEESLQWAGPQPPQPPQPLAGVAKYDLENMKKTAVPFRQTAQ  
GAQVPRTHSHVQPPQPPQPLAYKQRLQKQNDIASPLFPQGWAGGHYSKCT  
APTAQPHDPLTASSLAPQORVQNLHAYOSGRSLYDQOQOQOQOQOQALQSRH  
AQETHYLNKAKVHQOQOQYQCPDAVTRTEQYQTFSSSHSPASVGRSPSY  
SSTPSPLMNLNENPISQOPLGTGAFAGTIDSHHEPMLNPSPTDATSSPADLTALT  
SOVENISNTVQQLLSKAAVPQKKNLVSRTPEQHKMSQHCSPGSGYSPAGCTPL  
SEPPSTQSTHAPPEADYLSGEDPLERSFLYCNQAGSPARVNSKAPESYLG  
CSVTSDDMSKSDSFSQSLHGSPLDSFKFVAGERCDFRLLSALAQEDLASELIG  
LOEAIKCEADKAWAEAPSLKSSPPFLENHSAICLDSVAKSAWPRGPEPELPSL  
QLDKGNKDKSPGLPDSVAFAPDPKKTGTLPSFGTKPLGVAPDPTTAAPDF  
PDITAAASGANSFAPWENLGDACPRWGLHGLTKGLGGGKADGSLKSDTHEA  
SACLGFEQDFPPEKVASLPDGFQKEVEGVGKVEAGLQCEYAKADRWLEDSHCC  
STADGDLFLPPTSRKDELEAESEYSLCELLGSPQRPQMQLPSFKAPLICKEE

mRNA

gene

exon

CDS

BASE COUNT 4855 a 7314 c 6823 g 4848 t 3 others  
ORIGIN

alignment\_scores:  
Quality: 9115.00 Length: 1759  
Ratio: 5.209 Gaps: 4  
Percent Similarity: 99.488 Percent Identity: 99.375

alignment\_block:  
US-09-508-821B-6 x HSA271791 ..  
Align, seg 1/1 to: HSA271791 from: 1 to: 23843  
1 MetGlnSerPheArgGluArgCysGlyPheHisGlyLysGlnGlnAsnTy 17  
|||||  
9590. ATGCAGTCTTTTCGAGAAAGTGTGGTTTCATGCAACACAGACTA 9639  
|||||  
17 rGlnGluThrSerArgLeuGluAsnTyrArgGlnProS 34  
|||||

9640 CCAGCAGACCTGCGAGGAACATCAAGCCTAGAGAAATTACAGGACGCGA 9669  
34 erGlnAlaGlyLeuSerCysAspArgGlnArgLeuLeuAlaIAspTyr 50  
9690 GTCAAGCGCGGCTTAAGCTGAGACCGGCGACCGCTGCTGCGAAAGGACTAT 9739  
51 TyrAsnProGlnProTyrProSerTyrGlnGlyValAlaGlyThrProse 67  
9740 TATAACCGCGAGCTTACCGAGACTATGAGGTGGCGCTGGCAGCGCTTC 9789  
67 rGlyThrAlaAlaAlaValAlaAlaAspLysTyrHisArgGlySerLysA 84  
9790 TGGCACTGCAAGCGCGGTGGCGCGCAGACATACCAACGAGCAGACAGG 9839  
84 IalaLeuProThrGlnGlnGlyLeuGlnGlyArgProAlaIAspProGlyTyr 100  
9840 CCTGCCACACAGCAGCGCTGAGAGGAGCGCGCTTCCTGCTGCTGCTG 9889  
101 GlyValGlnAspSerSerProTyrProGlyArgTyrAlaGlnGlyLeu 117  
9890 GCGCTCCAGGACGACGACCGCTTACCGAGCGCTGCTGCTGCTGAGAG 9939  
117 rLeuGlnAlaTyrGlyAlaProGlnProProProProGlnProGlnProL 134  
9940 CCTTCAGGCTTGGGGGGCGCCACAGCAGCAGCAGCAGCAGCAGCAGCAG 9989  
134 euProAlaGlyValAlaLysTyrAspGlnAsnLeuMetLysGlyThrAla 150  
9990 TACCTGCAAGGCGGTGGCAAGTATGATGAACTGTGTAATAAAGCAGCA 10039  
151 ValProProSerArgGlnTyrAlaGlnGlyValAlaGlnValProPheAr 167  
10040 GTGCCCCCGACAGGACGATGACAGAGGCGCGCCAGGCGCTTCG 10089  
167 gThrHisSerLeuHisValGlnGlnProProProProGlnGlnProLeuA 184  
10090 GACTCACTCCCTGACGTCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 10139  
184 IatyrProLysLeuGlnArgGlnLysLeuGlnAsnAspIleHisSerPro 200  
10140 CAAACCCAGGCTCCAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 10189  
201 LeuProPheProGlnGlnGlyThrHisPheProGlnHisSerGlnSerPhePr 217  
10190 CTGCCCCCTTCCCCAGGCTACCACTTCTCTAGCAATTCACCACTCTTCCC 10239  
217 oThrHisSerThrTyrHisSerSerSerValGlnGlnGlyGlyGlnGlnAla 234  
10240 CACCTCTGCACTACTCTCTCTGCTGCAAGGAGGAGGAGGAGGAGGAGG 10288  
234 IAspSerTyrLysSerCysThrAlaProThrAlaGlnProHisAspArgPro 250  
10289 ACTCTATAGAGTGGACAGCAGCGACTGCCAGCCCCCAAGACAGGCGG 10338  
251 LeuThrAlaSerSerSerLeuAlaProGlyGlnArgValGlnAsnLeuH 267  
10339 CTGACTGCGAGCTCCAGCTTGGCGCGCGGAGCGGCTCCAGAACTCTCA 10388  
267 sAlaTyrGlnSerGlyArgLeuSerTyrAspGlnGlnGlnGlnGlnGlnG 284  
10389 TGGCTACAGTGGGGCGGCTCAGCTATGACACAGCAGCAGCAGCAGCAGC 10438  
284 LnglnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300  
10439 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 10488  
301 GlnThrLeuHisTyrGlnAsnLeuAlaLysTyrGlnHisTyrGlyGlnG 317  
10489 GAACCTCCATTTACCAAAACCTCGCCAAAGTATCAGCAGCAGCAGCAGCAG 10538  
317 nGlyGlnGlyTyrCysGlnProAspAlaAlaValArgThrProGlnGlnT 334  
10539 AGCCAGGAGGCTACTGCGAGCGGAGCAGCAGCTCGGAGCCAGAGCACT 10588  
334 yTyrGlnHisPheSerProSerSerSerHisSerProAlaArgSerVal 350  
10589 ACTACAGACCTTCAAGCCCCAGCTCCAGCAGCAGCAGCAGCAGCAGCAGC 10638  
351 GlyArgSerProSerTyrSerSerThrProSerProLeuMetProAsnLe 367  
10639 GAGCGCTACCTTCTACAGTTCCACACCTGCGCGCTGAGAGCAAACT 10688  
367 uGlnAsnPheProTyrSerGlnGlnProLeuSerThrGlyAlaIAspProA 384  
10689 GAGAACTTCTCCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 10738  
384 IaglyTLeuThrAspHisSerHisPheMetProLeuLeuAsnProSerPro 400  
10739 CAGGATCATGACACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 10788  
401 ThrAspAlaThrSerSerValAspThrAlaGlyAspCysLysProLe 417  
10789 ACAGATGCCACCACTCTGTGACACCCAGGCTGGCACTGCAAGCCCT 10838  
417 uGlnLysAspLysLeuProGlnAsnLeuLeuSerAspLeuSerLeuGln 434  
10839 TCAGAAAGACAAAGCTCCCTGAGAACCTGCTGCAATCTCAAGCTGCA 10887  
434 erLeuThrAlaLeuThrLeuGlnValGlnAsnLieserAsnThrValGln 450  
10888 C.CTCAAGCGCGTCACTCAAGGTGGAGAACATCTCAACACCTCCAG 10937  
451 GlnLeuLeuLeuSerLysAlaAlaValProGlnLysGlyValLysAs 467  
10938 CAGCTCTCTCTCCAGGCTGCTGCTGCGCGCAGAAAGAGTGTCAAA 10987  
467 nLeuValSerArgThrProGlnGlnHisLysSerGlnHisCysSerProG 484  
10988 CCGTGTGTCAGAGACCCAGCAGCAGCATAAAGCAGCAGCAGCAGCAGCAG 11037  
484 IuglySerGlyTyrSerAlaGlnProAlaGlyThrProLeuSerGlnPro 500  
11038 AAGGACCGGCTACACACCGCAGCGGCGCAGCAGCAGCAGCAGCAGCAG 11087  
501 ProSerSerThrProGlnSerThrHisAlaGlnProGlnGlnAlaAspTyr 517  
11088 CCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 11137  
517 rLeuSerGlySerGlnAspProLeuGlnArgSerPheLeuTyrCysAsnG 534  
11138 CCGTGAAGGCTCCGAGGAGCCACAGGAGGAGCAGCTTCTCTACTGCAAC 11187  
534 IAlaIArgGlySerProAlaArgValAsnSerHisSerLysAlaLysPro 550  
11188 AGGCGCTGGCAGCCCTGCGAGGCTCAACAGCAGCAGCAGCAGCAGCAG 11237  
551 GluSerValSerThrCysSerValThrSerProAspAspMetSerThrL 567  
11238 GAGTCCGCTGTCACCTGTTCTGTGACTCTCTGAGCAGCAGCAGCAGCAG 11287  
567 sSerAspAspSerPheGlnSerLeuHisGlySerLeuProLeuAspSerP 584  
11288 ACTGTAGACTCTCTCAAGCCTACAGCAGCAGCAGCAGCAGCAGCAGCAG 11337  
584 hSerLysPheValAlaGlyGlnArgAspCysArgLysLeuLeuLeuSer 600  
11338 TCTCCAACTTCTGTGGGGGTGAGGGGAGCTGCGGGGCTCTCTGAG 11387  
601 AlaLeuAlaGlnGlnAspLeuAlaSerGlnIleLeuGlyLeuGlnGln 617  
11388 GCGCTGGCAGCAGGAGGAGCTGAGCTCCGAATCTCTGGGCTGCGAGAGG 11437  
617 aIleGlyGlnLysAlaAspLysAlaTyrAlaIAspAlaIAspSerLeuVal 634  
11438 CATGGTGAAGAGCGAGCAAAAGCTTGGGCTGAGAGCAGCAGCAGCTGAT 11487

634 ysAspSerSerLysProProPheSerLeuGluAsnHisSerAlaCysLeu 650  
11488 AGGACACAGCAGAACCCCTTCTGCTGGAGAACACACAGCGCTGCGCTG 11537  
651 aspSerValAlaLysSerAlaLysProArgProGlyLeuProGluAlaLe 667  
11538 GACTCTGTGGCCAAAGAGTGGTGGCCCGGCTGGGAGACCCGAGGCCCT 11587  
667 pRoAspSerLeuGlnLeuAspLysGlyGlyAsnAlaLysAspPheSerP 684  
11588 GCGCGACTCTTGCAGCTGGCAGGAGGCGGCAATGCAAGAGACTTCAGCC 11637  
684 roGlyLeuPheGluAspProSerValAlaPheAlaThrProAspProLys 700  
11638 CAGGGCTGTTTGAAGACCTTCCGTGGCTTCGCTACGCTGACCCCAA 11687  
701 LysThrThrGlyProLeuSerPheGlyThrLysProThrLeuGlyValPr 717  
11688 AAGCAACTGGTCTCTCTCTCTGCTGACCAAGCCACCCCTTGGGGTTC 11737  
717 oAlaProAspProThrThrAlaAlaPheAspCysPheProAspThrThrA 734  
11738 TGCTCCAGACCCCTACTACAGACCTTTGACTGTTTCCGGACACACCG 11787  
734 lAlaSerSerAlaAspSerAlaAsnProPheAlaThrProGluGluAsn 750  
11788 CTGCCAGCTCAGCGGACAGGCGCAACCCCTTGGCTGGCCAGAGGAAAC 11837  
751 LeuGlyAsnAlaCysProArgTgGlyLeuHisProGlyGluLeuThrLy 767  
11838 CTGGGGAGTGTTCCTCCAGTGGGATGTCACCTGGGAGCTTACCAA 11887  
767 scLyLeuGluGlnGlyLysAlaSerAspGlyLysLeuLysGlyAspT 784  
11888 GGGCTCGAGCAGGCTGGGAGGCTTCAGATGGCTCAGCAAGGGGACA 11937  
784 hrHisGluAlaSerAlaCysLeuGlyPheGlnGluAspProGly 800  
11938 CCCATGAGCTTGGCTTGGCTTGGCTTCCAGGAGGAGGACCCCTTGGG 11987  
801 GluLysValAlaSerLeuProGlyAspPheLysGlnGluValGlyL 817  
11988 GAGAAAGTGGCTTGGCTGGCTGGCTTCAAGCAGGAGGAGTGGGTGG 12037  
817 yValLysGluAlaGlyLeuGluGlnCysProGluValAlaLysA 834  
12038 GTGAGAGGAGGAGGAGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 12087  
834 laAspArgTrpLeuGluAspSerArgHisCysSerThrAlaAspPhe 850  
12088 CTGACCGGTGGCTGGAGGACAGCCGCACTGCTGTCACCGCGACTTC 12137  
851 GlyAspLeuProLeuProProThrSerArgLysGluAspLeuGluAl 867  
12138 TGAGGAGGAGTACTCTCTCTCTGAGCTCTGGGACGCGCGAGCAGA 12237  
884 tProGlyMetGlnAspProLeuSerProLysAlaProLeuLeuCysThr 900  
12238 GCGCTGGCAGCAGGAGCCGCTGTCACCAAGGCGCCACTCATCTGCACC 12287  
901 LysGluGluValGluGluValLeuAspSerLysAlaGlyTrpGlySerPr 917  
12288 AAGAGAGAGTGGAGAGGCTGGAGCTCCAGAGCCGCGCTGGGCTTCC 12337  
917 oCysHisLeuSerGlyGluSerValLeuLeuLeuProThrValGlyT 934  
12338 GTGCCACCTCTCAGGAGAGTCCGCTCATCTGCTGGGCGCTACAGTGGCA 12387  
934 hrGluSerLysValGlnSerTrpPheGluSerSerLeuSerHisMetLys 950

12388 CCGAGTCAAAGTCCAGAGCTGTTGAGTCCCTCTGTCACACATGAAG 12437  
931 ProGlyGluGluGlyProAspGlyGluArgAlaProGlyAspSerThrTh 967  
12438 CAGGTGAGAGGGGCTGATGGGAGCGAGCTCCAGGGGATTCCACCAC 12487  
967 rSerAlaSerLeuAlaGlnLysProAsnLysProAlaValProGluA 984  
12488 CTGGACGCCCTCTCTGGCCAGAGCCCAACAGAGCTGCTGTGCCGAGG 12537  
984 laProLeuAlaLysLysGluProValProArgGlyLysSerLeuArgSer 1000  
12538 CGCCCATCGCAAGAGAGAGCTGTGCCAGGGGCAAAAGCTTACGGAGC 12587  
1001 ArgArgValHisArgGlyLeuProGluAlaGluAspSerProCysArgAl 1017  
12588 CGTCGGGTGCACCGGGGGCTGCCAGCGAGGAGCTCCCATCGAGGCG 12637  
1017 aProValLeuProLysAspLeuLeuLeuProGluSerCysThrGlyProP 1034  
12638 ACCAGTGTGCCCAGAGAGCTTGTCTCCCTGATCTGACACAGGGGCC 12687  
1034 roGlnGlyGlnMetGluGlyAlaGlyAlaProGlyArgGlyAlaSerGlu 1050  
12688 CCAGGGACAGATGGAAGGGGCTGGAGCCCGCAGCGGGGGGCTCGGAA 12737  
1051 GlyLeuProArgMetCysThrArgSerLeuThrAlaLeuSerGluProAr 1067  
12738 GGGCTCCCGAGAGTGTACTGCTTCTCACGGCCCTGAGTGAGCCCG 12787  
1067 gThrProGlyProProGlyLeuThrThrProAlaProProAspLysL 1084  
12788 CAGCGCCGAGCCCGCAGGCTGACACACCCCTGACCCCGCAGCAAAAC 12837  
1084 euGlyGlyLysGlnArgAlaAlaPheLysSerGlyLysArgValGlyLys 1100  
12838 TGGGGGCAACGACGAGCGGCTTCAGTGGGGCAAGCGGGTGGGAG 12887  
1104 ProSerProLysAlaAlaSerSerProSerAsnProAlaAlaLeuProVa 1117  
12888 CCCTCACCAAGGCTGCTCCAGCCCGCAGCAACCCGGCGGCTGCTGT 12937  
1117 laAspSerSerSerProMetGlySerLysThrLysGluThrAspSerP 1134  
12938 GGGCTCCGACAGCAGCGGCTGCGCTCCAGAGCAAGAGAGACAGCTCAC 12987  
1134 roSerThrProGlyLysAspGlnArgSerMetIleLeuArgSerArgThr 1150  
12988 CCAGCAGCGCTGGCAAGGACAGCGCTCCATGATCTCTCGGTCAAGCAC 13037  
1151 LysThrGlnGluLlePheHisSerLysArgArgProSerGluGlyAr 1167  
13038 AAAACCCAGAGAGATCTTCCACTCCCAAGCGGGAGGCGCTCTGAGGGCC 13087  
1167 gLeuProAsnCysArgAlaThrLysLysLeuLeuAspAsnSerHisLeuP 1184  
13088 GCTCCCGCACTGCGTGGCCACCAAGAGCTCTCGACACAGCCACTTGC 13137  
1184 roAlaThrPheLysValSerSerSerProGlnLysGluGlyArgValSer 1200  
13138 CCSCCACTTCAAGTCTCCAGCAGCCCGCAGAGGAGGAGGAGGAGTGA 13187  
1201 GlnArgAlaArgValProLysProGlyValaGlySerLysLeuSerAspAr 1217  
13188 CAGCGGGCAAGGGTCCCAACACTGCTGGTGGAGGAGGAGGAGGAGTGA 13237  
1217 gProLeuHisAlaLeuLysArgLysSerAlaPheMetaProValProT 1234  
13238 GCGCTCCATCGCTCAAAAGAGTGGGCTTCATGCGCGGCTGCCCA 13287  
1234 hrLysLysArgAsnLeuValLeuArgHisGlySerSerSerSerSerAsn 1250

13288 CCAAGAGCGGACCTGGCTTGGCGAGCGGACGACGACGACGAC 13337  
 1251 Alaser.Alametgylumetgylargargargylleuargvalprop 1267  
 13338 GCCAGTGGCAATGGGGGATGGGAAGGAGGAGGCGTGAAGGTTCCTCC 13387  
 1267 roProSerSerArggLyCysLeuLeuProArgLySerProProArg 1283  
 13388 CACCTCTTCAAGAGGATGTCTTCCCAAGAAAGCCACGACCAAGG 13437  
 1284 AlametAlaserLeuProGlnSerSerHisProArgArgProPrope 1300  
 13438 GCATGGCGAGCGCTGGCCAAAGCTCCACCCCGGAGGCCCGATGCC 13487  
 1300 eufProGlnAlaArgLeuSer.AlaAlaPheGlnGlyAlaMetLysThr 1316  
 13488 TGCCCTCAAGCTGGCTCTCGGGGAGCTTCAGGGGGCCATGAAGCAA 13537  
 1316 sValLeuProProArgLyArgGlyLeuLysLeuGlnAlaLeuValG 1333  
 13538 GGTGTGCGCCCGGAGGAGGCGGGGGCTGAAGCTGGAAGCCATCGTGC 13587  
 1333 InLysLeuThrSerProSerLeuLysLysPheAlaCysLysAlaProGly 1349  
 13588 AGAAGATCACCTCGCGCAGCTCCAGAGTTGGCATGTAAAGCGCAAGG 13637  
 1350 AlaSerProGlyAsnProLeuSerProSerLeuSerAspLysAspArg 1366  
 13638 GCGCTCTCTGGTAACTCTGAGACCCATCTTCGAGCAAAAGACCTGG 13687  
 1366 yLeuLysGlyAlaGlyLysSerProValGlyValGlnGlnGlyLeuVal 1383  
 13688 GCTCAAGGGTGTCTGGGGGAGCCAGCCAGTGGGGGTGAAGAGCCGTGAA 13737  
 1383 snValGlyThrGlyGlnLysLeuProThrSerGlyAlaAspProLeuG 1399  
 13738 ATGTGGGACCGGGGAGAGAGCTCCCACTTGTGGGCTATCGGTATGC 13787  
 1400 ArgAsnProThrAspArgSerLeuLysGlyLysLeuMetLysSerLys 1416  
 13788 AGAATCCAAACCAACAGATCTTAAAGGCAACTCATGAAGTAAGAA 13837  
 1416 sLeuSerSerThrAspCysPheLysThrGlnAlaPheThrSerProG 1433  
 13838 ACTGTCTTCTACTGACTGTTTCAAAACGAGGCTTCACATCCCGGAG 14887  
 1433 LeuLeuGlnProGlyGlyThrAlaLeuAlaProLysLysArgSerArg 1449  
 13888 CCTGTGACCTTGGGGGAGCTGGCTGGGCTTAAGAGAGGAGCCGAA 13937  
 1450 GlyArgAlaGlyAla.HisGlyLeuSerLysGlyProLeuGlnLysArg 1466  
 13938 GGCGGGGAGGAGGCGCCNATGACCTCCAAAGGCGCGCGAGAGGCG 13987  
 1466 roTyLeuGlyProAlaLeuLeuLeuThrProArgAspArgAlaSerGly 1482  
 13988 CCRNCTTGGCCCGGCTCTCTCTGACTCCCGAGAGCGGGCCAGTGG 14037  
 1483 ThrGlnGlyAlaSerGlnAspAsnSerGlyGlyGlyLysLysPro 1499  
 14038 ACACAAAGGGGCGGAGTGAAGCAACTGTGTGAGAGGAGCAAGACCAA 14087  
 1499 sMetGlnGlnLeuGlyProAlaSerGlnProProGlnGlnArgProCys 1516  
 14088 GATGGAGAGCTGGGCTGGCTCCACCGCCCGGAGGCGAGCCCTGGCC 14137  
 1516 InProGlnThrArgAlaGlnLysGlnProGlnHisThrAsnLysSer 1532  
 14138 AGCCCGACAGAAAGGACAGAAACAGCCAGGCGACCACTACAGAGAGC 14187  
 1533 TyrSerLysArgLysArgLeuThrArgGlyArgAlaLysAsnThrThr 1549  
 14188 TATTCCAGCGGAGCGCTCACTGCGGGGCGGCGCAAGAACCACTCTC 14237

1549 rSerProCysLysGlyArgAlaLysArgArgGlnGlnGlnValLeu 1566  
 14238 TTAACCTCTGAAGGGGGGTGCCAAGCAGACAGACAGACAGAGGTCTCC 14287  
 1566 roLeuAspProAlaGluProGlnLysArgLeuLysThrLysSerCys 1582  
 14288 CCTGTGATCCCGAGAGCGCTGAATCCGCTCAATATTAATTCCTTTC 14337  
 1583 LysArgLeuArgSerAspSerArgThrProAlaPheSerProPheVal 1599  
 14338 AAGCGCTGAGGTGACAGACCGAGACCCCGCTTCTCAACCTTGTGGG 14387  
 1599 gValGlnLysArgAspAlaPheThrThrLysCysThrValAlaAsnSer 1616  
 14388 GGTGAGAGAGCGAGACGCTTACACACCATATCATCTGTCTACATCCC 14437  
 1616 roGlyLysAlaProLysProHisArgLysProSerSerAlaSerSer 1632  
 14438 CTGGAGATGCGCCAGCCCGACAGAGAGCTTCTCTCTGCTCTCTCT 14487  
 1633 SerSerSerSerSerSerPheSerLeuAspAlaAlaGlyAlaSerLeu 1649  
 14488 TTCTCATCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 14537  
 1649 arThrLeuProGlyLysThrLeuGlnProArgProSerLeuProLeu 1666  
 14598 CACACTCTCTGAGAGCTTCATCTGACAGCGCGGCTCTCTCTCTCT 14587  
 1666 erSerThrMetHisLeuGlyProValAlaSerLysAlaLeuSerThr 1682  
 14588 CCGCACAGATGACATTTGGGGGCTGTGGTTCCAAAGCCCTGATGACT 14637  
 1683 CysLeuValCysCysLeuGlnAsnProAlaAsnLysAspLeu 1699  
 14638 TGCTTGTGTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 14687  
 1699 yAspLeuCysGlyProTyrTyrProGlnHisCysLeuProLysLys 1716  
 14688 GAGCTCTGTGGGCTCTACTACCTGGAACCTCTCTCTCTCTCTCTCT 14737  
 1716 roLysLeuLysGlnLysValArgProGlnGlyThrCysGlnGlnAla 1732  
 14738 CAATCTCAAGAGAGAGTGGCGGCGAGAGGACCTGAGAGAGGCTGT 14787  
 1733 LeuProLeuGlnArgThrLeuLysGlyProGlnCysAlaAlaAlaThr 1749  
 14788 CTCCCGCTTGAAGAGACCTCAAGAGTCCGAGTGGAGCTGCGGCAC 14837  
 1749 rAlaGlyLysProProArg 1755  
 14838 TGCCGGGAGAGCCCGCAGG 14856  
 seq\_name: gp\_pat:AX024201  
 seq\_documentation\_block: 5395 bp DNA PAT 15-SEP-2000  
 LOCUS AX024201 DEFINITION Sequence 1 from Patent WO0008143.  
 ACCESSION AX024201 VERSION AX024201.1 GI:10184512  
 KEYWORDS  
 SOURCE human;  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS Poustika,A., Lehrach,H., Radeflof,U. and Seranekski,P.  
 TITLES Gene isolated on the short arm of human chromosome 17  
 JOURNAL Patent: WO 0008143-A, 1 17-FEB-2000.  
 DEUTSCHES KREBSFORSCH (DE); POSTSTA ANNMARIE (DE); LEHRACH HANS  
 (DE); RADELLOF UWE (DE); SERANESKI PETER (DE); MAX PLANCK INST  
 FUER MOLEKULAR (DE)  
 FEATURES  
 location/Qualifiers



```
source 1. 5395
/organism="Homo sapiens"
/db_xref="taxon:9606"
CDS 1. >5394
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC08943.1"
/db_xref="GI:10184513"
/translation="MWSFRERCGHGKQNYQQTSTQSRLENYRQPSOAGLSCDRQR
LAKDYINQPSYSGGAGTSGTAAVAADKIRHGSGKALPTQGLQGRAPPGYGV
QSSPTPGYAGAESIQAWGAPPPQPPQPLQAGVAKYDENLAKTAVTPPSQAEG
GAQVPRSHLQVPPQPPQPLQAGVAKYDENLAKTAVTPPSQAEG
APTAQHPDRLTASSLAGQVOMLHAGSLYSYDQOQQOQQOQQOQQOQQOQQOQQO
AQETHYONLAKTGHYQOQOQVQCPDAVATPEYQYQFSSSHSPASVGRSPSY
SSTPSPLNLENPFTSQPLSTGAFAGTDSHSPHLLNPSPTDATSPADLTALT
SOVENISNTVOQLLLSKAAVPOKKGVKNLVERTPQEQKSHQCSPEGSGYSAEPAGTPL
SEPPSTQSTHAPQADYLSGSDPLERSFLYCMOARGSPARVNSNRKAEPSYST
CVSTSPDMSTKSDSGLSHSLDLSFSKVFAGERCDFRLLSALAQBDLASEILG
LQEAIGERADKAWAEPALVSDSSPPSLNHSACLDSDAKSAMPFRPEALPDSL
QLDKGNADKSPGLRDSVAFADPKTGTPLSGFKPTLGVPAAPDPTTAADF
PDTAASSNDSANPWPENLGDACPEWGLHPGELTKGLEOGKASDGLSKDQTHEA
SACLGQEDPPGKVASLPDFKQEVGVGVKHEAGLLQCEYAKAKADRWLESRHCC
STADFQDLPLPTSRKEDLAEETYSLLCHLSPQSPQPMQDPLSKAPLICTKEE
VEEVLDSKAGWSPCHLSGESVILLGPTVGTESKVSQSWFESSLSHMKPGBEDDGERA
LPXDTLLPESCTPGQOMEGAGAPGASGLPRMCTSLTALSEPRTPGPGLTAT
PAPDKLGKQRAAFKSGKYGKPSKASSPSPNPAALPVASDSPMGSKTKTWDSPS
TPGDKQRSLKSRKTKQETLFRHSRRRSEGLPNCRAKTKLLDNLHSLPSTKVSPP
QKGRYSORARVPKPGAGSLKSLDRHAKRKSAPMAPVPTKRLNLYLRSSSSSNA
SGMGKGKEERPEGSPLFLKMSPKKAKPTKNGEPATKLPPETPDACLKLSRAA
FOGAKTKVLPKRGRLKLEAVIKITSPSLKACFAGASPGNPLSPSLSDRQ
LKGAGSPVGVVEGLVNGVGTQPLTSGADPLCRNPTNRSKGLMNRASQTQASD
NSGSGGKPKMBELGLASQPEGRPCOPTRAKQKOPGHTNYSYKRRKLRGRKANT
TSPCKRARRKARQOVLDPDAPRILKYLSSCKELSDSRTFAFSPVVRKEDDA
PLSTVHLGVPVSGAPKHPKSPSSSSSSSSSSSSSSSLDAAGASLALPGLSILQPRSL
PLSTVHLGVPVSGAPKHPKSPSSSSSSSSSSSSSSSLDAAGASLALPGLSILQPRSL
VRPEGTCEASLPLERTLKPECAAAATAGRPDPADPAKQGLPTARSGLRKL
QSCYCCDGEADGGEAAAPKQHWHEACVAVTGVTVLAVKGLFLOAMKVAVDHNG
GTVLALAPDFSLPLGLRFASLFQGPSNCDPVLATSTPSSNSRCVPAAKKPGPLGAAT
KDASTPTTRVPM"
```

```
BASE COUNT 1117 a 1877 c 1595 g 806 t
ORIGIN
alignment_scores:
Quality: 8479.00 Length: 1758
Ratio: 5.189 Gaps: 6
Percent Similarity: 92.947 Percent Identity: 92.776
alignment_block:
US-09-508-821b-6 x AX024201
Align seg 1/1 to: AX024201 from: 1 to: 5395
1 MetGlnSerPheArgGluArgCysGlyPheHisGlyLysGlnInAsuTy 17
|||||
1 ATCCAGTCTTTTCGAGAAAGGTGTGGTTCCATGGCGCAACACAGAACTA 50
17 rGlnGlnThrSerGlnGlnThrSerArgLeuGluAsnTyArgGlnProS 34
|||||
51 CCAGCAGACCTCGCAGGAAACATCAGCGCTAGAGAAATTACAGCAGCCGA 100
34 erGlnAlaGlyLeuSerCysAspArgGlnArgLeuLeuAlaLysAspTy 50
|||||
101 GTCAGCGCGGGCTAAGCTGCGCGGAGCGGCTGCTCGCCAGGACUAT 150
51 TyrAsnProGlnProTyProSerTyArgGlyGlyAlaGlyThrProse 67
|||||
151 TATACCCCGACGCTTACCCGACCTATAGAGGTGGCGCTGGCAGCCCTC 200
67 rGlyThrAlaAlaValAlaAlaAspLysTyThrHisArgGlySerLysA 84
|||||
201 TGGCACTGCGAGCGCGGTGGCCGCGCAAGTACACCGAGGCGCAAGG 250
```

```
84 laLeuProthrGlnGlnGlnGlyLeuGlnGlyArgProAlaPheProGlyTy 100
|||||
251 GCCTGCCACACAGCAAGGCTGTCAGGGAGGCGGCTTCCCTGGCTAC 300
101 GlyValGlnAspSerSerProTyProGlyArgTyArgAlaGlyGluGluSe 117
|||||
301 GCCTGCCAGCAGCAGCAGCCCTACCCAGGCGGCTATGCTGGTGGAGGAG 350
117 rGlnGlnAlaTTPGlyAlaProGlnProProProProProGlnProGlnProL 134
|||||
351 CTTTCAGGCTGGGGGGGGCCACAGCCACCCACCCAGCCAGCGCGCAC 400
134 euProAlaGlyValAlaLysTyArgGluAsnLeuMetLysLysThrAla 150
|||||
401 TACCTGCGAGGCTGCCAAGTATGATGAGAACTTGTATGAAAAAGACAGCA 450
151 ValProProSerArgGlnTyArgAlaGlnGlnGlyAlaGlnValProPheAr 167
|||||
451 GTGCCCCCAGCAGCAGGATGTCAGAGCAGGCGGCCAGCGCCAGCGCCCTGG 500
167 gThrHisSerLeuHisValGlnGlnProProProProProGlnGlnProLeuA 184
|||||
501 GACTCACTCCCTGCAGCTCCAGCAGCCAGCGCGGCCAGCGCCAGCGCCCTGG 550
184 laTyProLysLeuGlnArgGlnLysLeuGlnLysLeuGlnLysAspIleAlaSerPro 200
|||||
551 CATACCCCAAGCTCCAAAGGCAAGAGCTGCAACACACATTCGCTCCCTCC 600
201 LeuProPheProGlnGlyThrHisPheProGlnHisSerGlnSerPhePr 217
|||||
601 CTGCCCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 612
217 oThrSerSerThrTySerSerSerValGlnGlnGlyGlyGlnGlnGlnAlaH 234
|||||
613 .....CAGGGGTGGTGGCGAGGGGCGCC 634
234 isSerTyLysSerCysThrAlaProThrAlaGlnProHisAspArgPro 250
|||||
635 ACTCTATAAGAGTGTGCACAGCAGCGGCTGCGCCAGCGCCATGACAGGCGG 684
251 LeuThrAlaSerSerSerLeuAlaProGlyGlnArgValGlnAsnLeuHi 267
|||||
685 CTGACTGCCAGCTCCAGCTTGCGCGCGGCGAGCGGGTCCAGAACTTCA 734
267 saLaTyGlnSerGlyArgLeuSerTyArgGlnGlnGlnGlnGlnGlnGln 284
|||||
735 TGCTTACCACTGCGGCGGCTCAGCTATGACGACGACGACGACGACGACG 784
284 lnglnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300
785 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 834
301 GluThrLeuHisTyArgGlnAsnLeuAlaLysTyArgGlnGlnGlnGln 317
|||||
835 GAACCCCTCCATTACCAAAACCTCGCCAGTATCAGCACTACGGGCGACGA 884
317 nGlyGlnGlyTyCysGlnProAspAlaAlaValArgThrProGlnGlnT 334
|||||
885 AGGCCAGGCTACTCCAGCGCGAGCAGCGCTCGCGGACCCAGCAGCAGT 934
334 yTyArgGlnThrPheSerProSerSerHisSerHisSerProAlaArgSerVal 350
935 ACTACCAAGACCTTCAGCCCCAGCTCAGCCAGCAGCAGCAGCAGCAGCAG 984
351 GlyArgSerProSerTySerSerThrProSerProSerProLeuMetProAsnLe 367
|||||
985 GCGCGCTCAGCTTCTTCTACAGTTCACACCGCTGCGCTGATGCCAACT 1034
367 uGluAsnPheProTySerGlnGlnProLeuSerThrGlyAlaPheProA 384
1035 GGAGAACTTTCCTACAGCAGCAGCGCTCAGCAGCGGCGGCTCCCGG 1084
```



384 IagIYIlethrasphHisSerHisPheMetProLeuLeuAnProSerPro 400  
1085 CAGGATCACTGACCAAGCAAGCTTCACTGCCCCCTGCTCAATCCCTCCCA 1134  
401 ThrAspAlaThrSerSerValAspThrGlnAlaGlyAsnGlySerProle 417  
1135 ACGATGCCACCAAGCTCTCT ..... 1155  
417 uGlnLysAspLysLeuProGlnAsnLeuSerAspLeuSerLeuGlns 434  
1156 ..... GCAg 1159  
434 erLeuThrAlaLeuThrLeuGlnValGlnAsnLeSerAsnThrValGln 450  
1160 ACCTCAGCGCTGACCTCAAGGTGAGAACATCTCCAAACCGCTCCAG 1209  
451 GlnLeuLeuSerLysAlaAlaValProGlnLysGlyValLysAs 467  
1210 CAGCTGCTGCTCCAAAGGCTGCTGCTCCAGAAAGAAAGGTGCAAGAA 1259  
467 nLeuValSerArgThrProGlnGlnHisLysSerGlnHisGlySerPro 484  
1260 CCGCTGTCAGAGACCCAGACAGCAATAAAGCCAGCACTGCAAGCCCG 1309  
484 LuGlySerGlyTyrSerAlaGlnProAlaGlyThrProLeuSerGluPro 500  
1310 AAGGAGCGGCTACTAGCCGAGGAGCCGAGGACACCGCTGCAAGACCG 1359  
501 ProSerSerThrProGlnSerThrHisAlaGlnProGlnGlnAlaAsp 517  
1360 CCGAGCAGCAAGCAAGCTCCAGCATCGGAGCCGAGAGGAGGCGGCTA 1409  
517 rLeuSerGlySerGlnAspProLeuGlnLysSerPheLeuTyrCysAsn 534  
1410 CCTGAGCGGCTCCAGAGACCACTGAGCGAGCTTCTCTCACTCAACC 1459  
534 LnlAlaArgGlySerProAlaArgValAsnSerAsnSerLysAlaLysPro 550  
1460 AGGCGCGTGAGCCCTGCGAGGCTCAACAGCAACTCGAAGGCCAAGCC 1509  
551 GluSerValSerThrCysSerValThrSerProAspAspMetSerThr 567  
1510 GATCGGTGTCAACCTGTCTGTGACCTCTCTGACAGCATGTCCACCA 1559  
567 sSerAspAspSerPheGlnSerLeuHisGlySerLeuProLeuAspSer 584  
1560 ATCTGAGCACTCTTCCAGAGCTTACAGGAGGTCTGCGCTGACACT 1609  
584 hSerLysPheValAlaGlyLysArgAspCysProArgLeuLeuLeuSer 600  
1610 TCTCCAGTCTGCGCGGTGAGCGGAGCTGCGCGGCTGCTGCTCAGC 1659  
601 AlaLeuAlaGlnGlnAspLeuAlaSerGlnLeuGlnLeuGlnGln 617  
1660 GCCCTGCAAGAGAGACCTGCGCTCCAGATCTCTGGGCTGCAAGAGC 1709  
617 aLleGlyGlnLysAlaAspLysAlaThrAlaGlnAlaProSerLeuVal 634  
1710 CATCGGTGAGAAAGCCGCAAAAGCTTGGGCTTAAACCAACCGCTGTA 1759  
634 yAspSerSerLysProProPheSerLeuGlnAsnHisSerLysAlaCysLeu 650  
1760 AGGACAGCAAGACCACTTCTCGTGGAGAACACACACCGCTCCCTG 1809  
651 AspSerValAlaLysSerAlaThrProArgProGlyGlnProGlnAlaLe 667  
1810 GACTCTGTGGCCAAAGGTGGTGGCCCGGCTGGGAGCGGAGGCGCT 1859  
667 uProAspSerLeuGlnLeuAspLysGlyLysAlaLysAspPheSer 684  
1860 GCCGAGCTCTTCCAGCTGAGCAAGGCGGCAATGCTCAAGACTCAAGC 1909  
684 roGlyLeuPheGlnAspProSerValAlaPheAlaThrProAspProLys 700

1910 CAGGCTGTGTGAAGACCTTCCGAGCTTCCCTCAAGCTGACCAAAA 1959  
701 LysThrThrGlyProLeuSerPheGlyThrLysProThrLeuGlyValP 717  
1960 AAGACACTGTGCTCTCTCTTGTGACCAAGCCACCTGAGGCTTCC 2009  
717 oAlaProAspProThrAlaAlaPheAspCysPheProAspPheThr 734  
2010 TGCTCAGACCCCACTCAAGCAAGCTTGTGATGTTCCCGGACCAACCG 2059  
734 lAlaSerSerAlaAspSerAlaAsnProPheAlaThrProGlnGlnAsn 750  
2060 CTCCAGCTCAAGCAAGGACCAACCTTGTGCTGCGCAAGAGAAAAC 2109  
751 LeuGlyAspAlaCysProArgThrProGlyLeuHisProGlyGlnLeuThr 767  
2110 CTGGGGAGTCTGTCCAGGTGGGATTTGCAACCTGAGACTTACCA 2159  
767 sGlyLeuGlnGlnGlyLysAlaSerAspGlyLleSerLysGlyAsp 784  
2160 GGGCTGAGCAGGAGTGGAGAGCTCAAGATGATCATGCAAGAGGACA 2209  
784 hHisGlnAlaSerAlaCysLeuGlyPheGlnGlnGlnAspProGly 800  
2210 CCATGAGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2259  
801 GlnLysValAlaSerLeuProGlyAspPheLysGlnGlnValGlyG 817  
2260 GAGAGGTGGCTGCTTCCCGGGGACTTCAAGAGAGAGAGAGTGGTGG 2309  
817 yValLysGlnGlnAlaGlyLysLeuLeuLncProGlnValAlaLys 834  
2310 GGTGAAGAGAGAGAGTGGGCTGCTGAGAGTCCGAGAGTGGCCAAAG 2359  
834 lAspArgThrLeuGlnAspSerArgHisCysCysSerThrAlaAspPhe 850  
2360 CTGACCGGTGCTGAGAGACCGGAGCTGCTGCTGCTGCTGCTGCTG 2409  
851 GlyAspLeuProLeuLeuProThrSerArgLysGlnAspLeuGln 867  
2410 GGGAGCTCCCACTGTGCAACCAAGAGAGAGAGAGAGAGAGAGAG 2459  
867 aGlnGlnGlyTyrSerSerLeuGlyLeuLeuGlySerProGlnGln 884  
2460 TGAAG 2509  
884 rProGlyMetGlnAspProLeuSerProLysAlaProLeuLleCysThr 900  
2510 GGCCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2559  
901 LysGlnGlnValGlnGlnValLeuAspSerLysAlaGlyTyrGlySerP 917  
2560 AAG 2609  
917 oCysHisLeuSerGlyGlnSerValLleLeuLeuGlyProThrValGly 934  
2610 GTGCACTCTCAAGGAGAGTCCGCTATCTGCTGAGGAGAGAGAGAG 2659  
934 hGlnSerLysValGlnSerThrPheGlnSerSerLeuSerHisMetLys 950  
2660 CCGAGTCAAGAGTCAAGAGTGTGTGAGTCTCTGTGCAACATAGAG 2709  
951 ProGlyGlnGlnGlyProAspLysGlnArgAlaProGlyAspSerThr 967  
2710 CAGGTGAG 2759  
967 rSerAspAlaSerLeuAlaGlnLysProAsnLysProAlaValProGln 984  
2760 CTGAG 2809  
984 lArProLleAlaLysGlnProValProArgLysSerLeuArgSer 1000

2810 GCGCCATCGCAAGAGAGCTGTGCCAGGGGCAAGAGCTTACGGAGC 2859  
1001 ArgArgValHisArgGlyLeuProGluAlaGluAspSerProCysArgAl 1017  
2860 CGTCGGGGTGCACGGGGGGTCCCGAGGCGGAGGACTCCCGATGCGAGGC 2909  
1017 aProValLeuProLysAspLeuLeuProGluSerCysThrGlyProp 1034  
2910 ACCAGTGTGCCCAAGAGACCTTGTCTCCCTGATCTCTGACAGGGCCCC 2959  
1034 roGlnGlyGlnMetGlnGlyAlaGlyAlaProGlyArgGlyAlaSerGlu 1050  
2960 CCAAGGGACAGATGGAAGGGCTGGAAGCCCGCCAGGCGGGGGCTCGGAA 3009  
1051 GlyLeuProArgMetCysThrArgSerLeuThrAlaLeuSerGluProAr 1067  
3010 GGGCTCCCGAGAGTGTACTCTCTCTCAGCGGCTGAGTGGCCCG 3059  
1067 gThrProGlyProGlyLeuThrThrProAlaProProAspLysL 1084  
3060 CAGCGCCGAGACCCCGAGGCTGACACACCCCTGCACCCCGACACAAAC 3109  
1084 euGlyGlyGlnArgAlaAlaPheLysSerGlyLysArgValGlyLys 1100  
3110 TGGGGGCGACAGCGAGCGCCCTTCAAGTCGGGCAAGCGGTGGGGAAG 3159  
1101 ProSerProLysAlaAlaSerSerProSerAsnProAlaAlaLeuProVa 1117  
3160 CCGTCACCAAGCTGCCCTCCAGCCCGACACCCCGCGCGCTGCCGT 3209  
1117 AlaSerAspSerProMetGlySerLysThrLysThrAspSerP 1134  
3210 GCGCTCCGACAGCGCCGATGGGCTCCAGGACCAAGAGACAGACTCAC 3259  
1134 roSerThrProGlyLysAspGlnArgSerMetIleLeuArgSerArgThr 1150  
3260 CCAGCAGCGCTGGCAAGGACCGAGCGCTCCATCATCTTCGTCAGCGACC 3309  
1151 LysThrGlnGlnPheHisSerLysArgArgArgProSerGluGlyAr 1167  
3310 AAACCCAGAGAGATCTTCACTCAGCGGGGAGGCGCCCTCGAGGGCGG 3359  
1167 gLeuProAsnCysArgAlaThrLysLysLeuLeuAspAsnSerHisLeup 1184  
3360 GCTCCCAACTGCGCTGCCACCAAGAGCTCTCCGACACAGCCACTTGC 3409  
1184 roAlaThrPheLysValSerSerProGlnLysGluGlyArgValSer 1200  
3410 CCGCCCATTTCAAGGTCTCCAGAGCGCCCGCCAGAGGAGGCGAGGTGAGC 3459  
1201 GlnArgAlaArgValProLysProGlyValGlySerLysLeuSerAspAr 1217  
3460 CAGCGGGCAAGGTCCCAACCTGTGTCAGGACGACAGCTCTCTGACCG 3509  
1217 gProLeuHisAlaLeuLysArgLysSerAlaPheMetAlaProValProt 1234  
3510 GCGCTCCATGCGCTCAAGAGAGTCTGCGCTTCATGGCGCGGTCCCA 3559  
1234 hrLysLysArgAsnLeuValLeuArgHisGlySerSerSerSerAsn 1250  
3560 CCAAGAGGAGGAGTGTCTTCCGAGCGCCGACACAGCAGCAGAGCAAC 3609  
1251 AlaSerAlaMetGlyGlnMetGlyArgArgGlyLeuArgValProp 1267  
3610 GCGAGTGGCAATGGGGAGATGGGAAGGAGGAGGCGCTGAGGGTTCCC 3659  
1267 roProSerSerArgGlyCysLeuLeuProArgLysProSerProProArg 1283  
3660 CACCTCTTCAAGAGATGTCTTCTCCCAAGAAAGCCAGCCACCCAGG 3709  
1284 AlaMetAlaSerLeuProGlnSerSerHisProArgProProPhe.L 1300  
3710 GCATGGCGGAGCTGCCAAGAGCTCCCAACCCCGGAGACCCCGATGCC 3759

1300 euProGlnAlaArgLeuSerAlaAlaPheGlnGlyAlaMetLysThrLys 1316  
3760 TGCTCAAGCTCGCTCTCGGACAGCTTCCAGGGGGCCCAAGAGACCA 3809  
1316 sValLeuProProArgLysGlyArgGlyLeuLysLeuGluAlaIleValG 1333  
3810 GGTGCTGGCCACCCCGAAGGGCGGGGCTGAAGCTGGAAGCCATCGTGC 3859  
1333 LnyLysIleThrSerProSerLeuLysLysPheAlaCysLysAlaProGly 1349  
3860 AGAAGATCACTCCCGCCAGCCTCAGAGAGTCCCATGTAAAGCCCGAGG 3909  
1350 AlaSerProGlyAsnProLeuSerProSerLeuSerAspLysAspArgG 1366  
3910 GCCTCTCTCTGTAATCTCTGAGCCCATCTTCCGACAAAGACCGTGG 3959  
1366 YLeuLysGlyAlaGlyLysSerProValGlyValGlnGlyLeuVala 1383  
3960 GCTCAAGGGTCTTGGGGGAGCCCTAGTGGGGGTGAGAGAGCCCTGTAA 4009  
1383 snValGlyThrGlyGlnLysLeuProThrSerGlyAlaAspProLeuCys 1399  
4010 ATGTGGGACCGGGCAGAAAGCTCCCAACTTCTGGGCTCATCGTTATOC 4059  
1400 ArgAsnProThrAsnArgSerLeuLysGlyLysLeuMetAsnSerLysL 1416  
4060 AGAATCCACACCAACAGATCTTAAAGGCAAACTCATGAAC..... 4101  
1416 sLeuSerSerThrAspCysPheLysThrGluAlaPheThrSerProGlu 1433  
4101 ..... 4101  
1433 laLeuGlnProGlyThrAlaLeuAlaProLysLysArgSerArgLys 1449  
4101 ..... 4101  
1450 GlyArgAlaGlyAlaHisGlyLeuSerLysGlyProLeuGluLysArgPr 1466  
4101 ..... 4101  
1466 oTyLeuGlyProAlaLeuLeuThrProArgAspArgAlaSerGlyT 1483  
4102 .....AGGCCAGTGGCA 4114  
1483 hrGlnGlyAlaSerGluAspAsnSerGlyGlyLysLysProLys 1499  
4115 CACAAGGGCCAGTGGAGGACAACTCTGTTGGAGGAGGCAAGAGCCAAAG 4164  
1500 MetGluGluLeuGlyProAlaSerGlnProProGluGlyArgProCysG 1516  
4165 ATGGAGGAGCTGGGCTTGGCTTCCCGCCCGGGAGGCGAGGCGCTTGC 4214  
1516 nProGlnThrArgAlaGlnLysGlnProGlyHisThrAsnTySerSerT 1533  
4215 GCGCCAGACAAAGGCGACAGAAACAGCCAGGCCACCACTACAGCAGCT 4264  
1533 yrSerLysArgLysArgLeuThrArgGlyArgAlaLysAsnThrThrSer 1549  
4265 ATTCCAGCGAGGCGCTTCTACTCGGGGCGGGGCCAAGAACACCACTCT 4314  
1550 SerProCysLysGlyArgAlaLysArgArgGlnGlnGlnValLeuPr 1566  
4315 TCACCTCTTAAGGGCGTGGCAAGCAGACACAGCAGCAGCGTGTGCC 4364  
1566 oLeuAspProAlaGluProGluLeuLeuArgLeuLysTyIleSerSerCysL 1583  
4365 CTGTGATCCCGAGAGCTTGAATCCCGCTCAAGTACATTTCTCTTCCA 4414  
1583 ysArgLeuArgSerSerArgThrProAlaPheSerProPheValArg 1599  
4415 AGCGGCTGAGGTGAGAGCGGAGCCCGGCTTCTCACCTTCGTGGGG 4464







```
3710 GCATGGCGAGCGCTGCCACAAGCTCCACCCCGGAGACCCCGCATGCC 3759
1300 euProGlnAlaArgLeuSer. AlalaPheGlnGlnAlaMetLysThrLy 1316
|||||
3760 TGCTCAAGCTCGCTCTCGGGCAGCCTTCAGGGGGCCCATGAAGACCA 3809
1316 sValLeuProProArgLysGlyArgGlyLeuLysLeuGlnAlaIleValG 1333
|||||
3810 GGTGCTGCCACCGCCGGAAGCGCGGGCTGAAAGCTGGAAGCCATCGTGC 3859
1333 InLysIleThrSerProSerLeuLysLysPheAlaCysLysAlaProGly 1349
3860 AGAAGATCACTCGCCGAGCTCAAGAGTTCGATGTAAGCGCAGGG 3909
1350 AlaSerProGlyAsnProLeuSerProSerLeuSerAspLysAspArgL 1366
|||||
3910 GCCTCTCTGTAATCTCTGAGCCCATCCCTTTTCGACAAAGACCGTGG 3959
1366 yLeuLysGlyAlaGlyGlySerProValGlyValGlnGlnGlyLeuValA 1383
3960 GCTCAAGGGTGTGGGGCAGCCAGTGGGGGTGGAAGAGCGCTGGTAA 4009
1383 snValGlyThrGlyGlnLysLeuProThrSerGlyAlaAspProLeuCys 1399
4010 ATGTGGCCACCGGCGAGAGCTCCCAACTCTGGGGCTGATCGTTATGC 4059
1400 ArgAsnProThrAspSerLeuLysGlyLysLeuMetAsnSerLysLy 1416
|||||
4060 AGAAATCCAAACCAAGATCTTAAAGCAACATCATGAAC 4101
1416 sLeuSerSerThrAspCysPheLysThrGlnAlaPheThrSerProGluA 1433
4101 ..... 4101
1433 laLeuGlnProGlyGlyThrAlaLeuAlaProLysLysArgSerArgLys 1449
4101 ..... 4101
1450 GlyArgAlaGlyAlaHisClyLeuSerLysGlyProLeuGluLysArgPr 1466
4101 ..... 4101
1466 ofyLeuGlyProAlaLeuLeuThrProArgaspArgAlaSerGlyT 1483
4102 ..... AGGCCAGTGGCA 4114
1483 hrGlnGlyAlaSerGluAspAsnSerGlyGlyLysLysProLys 1499
4115 CACAAGGGCCAGTGGAGACACTCTGTGGAGGAGCAAGCAAGCAAG 4164
1500 MetGluGluLeuGlyProAlaSerGlnProProGluGlyArgProCysG 1516
4165 ATGGAGAGCTGGGCTGGCTCCAGCCCCCGGAGCGAGGCCCTGCCA 4214
1516 nProGlnThrArgAlaGlnLysGlnProGlyHisThrAsnTySerSerT 1533
4215 GCCCCAGACAGGCGCAGAAACACCCAGGCCACACCACTACAGCAGCT 4264
1533 TySerLysArgLysArgLeuThrArgGlyArgAlaLysAsnThrThrSer 1549
4265 ATTCCAAGCGGAGCGCTCACTCGGGCGCGGGCGCAAGCAACACCTCT 4314
1550 SerProCysLysGlyArgAlaLysArgArgGlnGlnGlnValLeuPr 1566
4315 TCACCTGTGAAGGGCGTGCAGAGCGGAGCAGCAGCAGCAGCTCTGCC 4364
1566 oLeuAspProAlaGluProGluIleArgLeuLysTyThrIleSerSerCysL 1583
4365 CCGTGAATCCCGCAGAGCTGAAATCCGCTCGCTCAATACATTTCTCTTCCA 4414
1583 ysrArgLeuArgSerAspSerArgThrProAlaPheSerProPheValArg 1599
4415 AGCGGCTGAGGTCACAGACCGGAGCCCGGCTTCTCACCTCTCGTGGG 4464
```

```
1606 ValGluLysArgAspAlaPheThrThrIleCysThrValValAsnSerPr 1616
|||||
4465 GTGGAGAGCGAGACGCGTTCAACCATATGCACTGTGTCAACTCCCC 4514
1616 oGlyAspAlaProLysProHisArgLysProSerSerAlaSerSerS 1633
4515 TGAGAGATGCCCAAGCCCAAGAGAGGCTTCTCTCTGCGCTCTCTCT 4564
1633 erSerSerSerSerPheSerLeuAspAlaAlaGlyAlaSerLeuAla 1649
4565 GCTATATCTGCTCTGCTCTCTCTGATGTCAGCGGGGCTCTCTGGCC 4614
1650 ThrLeuProGlyLysIleLeuGlnProArgProSerLeuProLeuSe 1666
4615 ACACCTCCCTGGAGGCTCCATCTGCAGCGCGGCTCTCTGCGCTCTC 4664
1666 rSerThrMetHisLeuGlyProValValSerLysLysAlaLeuSerThr 1683
4665 CTCACGATGCACCTTGGGCTGTGGTTTCCAAGGCTCTGAGTACCTCT 4714
1683 ysLeuValCysLysLeuCysGlnAsnProAlaAsnPheLysAspLeuGly 1699
4715 GCCTTGTGTGCTCTGCTCTGCCAAACCGGCCCACTTCAAGGACCTTGG 4764
1700 AspLeuCysGlyProTyrrProGluHisCysLeuProLysLysLysPr 1716
4755 GACCTCTGTGGGCTACTACCTGAACACTGCTCTCCCAAAAGAGGCC 4814
1716 oLysLeuLysGluLysValArgProGluGlyThrCysGluGluAlaSerL 1733
4815 AAACTCAAGAGAGGTGGCGGCAAGAGCACTGTGTGAGGAGGCTCGC 4864
1733 euProLeuGluArgThrLeuLysGlyProGluCysAlaAlaAlaThr 1749
4865 TGCGCTGTGAGAGACACTCAAGGTCCCGAGTGTGACGTGCGCCCACT 4914
1750 AlaGlyLysProProArg 1755
4915 GCGGGAGAGCCCGCCAGG 4932
seq_name: gb_pr:AB058723
seq_documentation_block:
LOCUS AB058723 5915 bp mRNA PRI 05-JUN-2001
DEFINITION Homo sapiens mRNA for KIAA1820 protein, partial cds.
ACCESSION AB058723
VERSION AB058723.1 GI:14017856
KEYWORDS
SOURCE:
ORGANISM Homo sapiens brain cDNA to mRNA, clone:hh01321.
REFERENCE
AUTHORS Nagase,T., Nakayama,M., Nakajima,D., Kikuno,R. and Ohara,O.
TITLE Prediction of the coding sequences of unidentified human genes. XX.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro
JOURNAL DNA Res. 8 (2), 85-95 (2001)
MEDLINE 21245130
REFERENCE 2 (bases 1 to 5915)
AUTHORS Ohara,O., Nagase,T. and Kikuno,R.
TITLE Direct Submission
JOURNAL Submitted (27-NAR-2001) Osamu Ohara, Kazusa DNA Research Institute,
Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba
292-0812, Japan (E-mail:cdna.info@kazusa.or.jp,
URL:http://www.kazusa.or.jp/huge, Tel:81-438-52-3913,
Fax:81-438-52-3914)
FEATURES
source
1. 5915
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hh01321"
```

gene  
/tissue\_type="brain"  
/note="vector:pb1uescriptII SK plus"  
458..5392  
/gene="K1A1820"  
CDS  
/gene="K1A1820"  
/note="Start codon is not identified."  
/codon\_start=1  
/product="K1A1820 protein"  
/protein\_id="BAB47449.1"  
/db\_xref="GI:14017857"

alignment\_scores:  
Quality: 8355.50 Length: 1741  
Ratio: 5.092 Gaps: 8  
Percent Similarity: 94.256 Percent Identity: 92.648  
align ment\_block:  
US-09-508-821b-6 x AB058723 ..  
Align seg 1/1 to: AB058723 from: 1 to: 5915  
1 MetGlnSerPheArgGlnArgGlyPheHisGlyLysGlnGlnAsnTy 17  
470 ATGCAAGCTTTTCGAGAAAGGTGGTTCCATGCGAAAGAAAGAACTA 519  
17 TGlNGlInTSeRgInGlnTInTSeArGleuGlnAsnTyArgGlnProS 34  
520 CCAAGCAACCTCGCAGAAATCATCAGCTAGAAATTAACAGCACCGA 569  
34 erGlnAlaGlyLeuSerCysAspArgGlnArgLeuLeuAlaLysAspTy 50  
570 GTCAGGCGCGGCTTAAGCTCGACCGGACGCGCTGCGCAAGAGACTAT 619  
51 TyrAspProGlnProTyrProSerTyrGlnGlyGlyAlaGlyThrProSe 67  
620 TATAACCCGAGCTTACCGAGCTATGAGGTGGCGCTGGCAGCGCCCTC 669  
67 rGlyThrAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 84  
670 TGGCACTGACAGCGCGGCTGGCGCGCGCAAGTACACCGAGAGAGCAAG 719

84 lAlaLeuProThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 100  
720 CCGTCCAGGCTTGGGGGGCCCGACAGCCACCGCCCGCCAGGCTTCCG 769  
101 GAlaValGlnAspSerSerProTyrProGlyArgTyrAlaGlyGlnGln 117  
770 GGGCTCAG 819  
117 rLeuGlnAlaTyrGlyAlaProGlnProProProProGlnProGlnPro 134  
820 CCGTCCAGGCTTGGGGGGCCCGACAGCCACCGCCCGCCAGGCTTCCG 869  
134 eudProAlaGlyAlaAlaGlyTyrAspGlnAsnLeuLeuLysTyrThrAla 150  
870 TACCTCAGAGGCTGCGCAAGTATGAGAACTTGAAGAAAGAGAGCA 919  
151 ValProProSerArgGlnTyrAlaGlnGlnGlnGlnGlnGlnGlnGln 167  
920 GTCGCCCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 969  
167 gThrHisSerLeuHisValGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 184  
970 GACTCAGCTCCCGACAGCGTCCAGAGAGAGAGAGAGAGAGAGAGAGAG 1019  
194 lArgProLysLeuGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 200  
1020 CATACCCCAAGCTCCAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1069  
201 LeuProPheProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 217  
1070 CTGCGCTCCCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1119  
217 oThrSerSerThrTyrSerSerSerValGlnGlnGlnGlnGlnGlnGln 234  
1120 CAGCTCAGCTCCCGACAGCGTCCAGAGAGAGAGAGAGAGAGAGAGAGAG 1169  
234 lSerSerTyrLysSerCysThrAlaProThrAlaGlnProHisAspArgPro 250  
1170 ACTCGCTTAAGAGATTGCACAGACAGAGAGAGAGAGAGAGAGAGAGAG 1219  
251 LeuThrAlaSerSerSerLeuAlaProGlnGlnGlnGlnGlnGlnGln 267  
1220 CTGAGCTCAGCTCCCGACAGCGTCCAGAGAGAGAGAGAGAGAGAGAGAG 1269  
267 sAlaTyrGlnSerGlyArgLeuSerTyrAspGlnGlnGlnGlnGlnGln 284  
1270 TCCCTACAGAGTCCCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1319  
284 lGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300  
1320 AGCAGCAG 1369  
301 GluThrLeuHisTyrGlnAsnLeuAlaLysTyrGlnHisTyrGlnGln 317  
1370 GAACCTCCATTAACAAACCTCGCAAGTATACACACACAGGAGAGCA 1419  
317 nGlyGlnGlnTyrCysGlnProAspAlaAlaValAlaArgThrProGlnGln 334  
1420 AGGCGAG 1469  
334 yTyrGlnThrPheSerProSerSerSerHisSerProAlaTyrSerVal 350  
1470 ACTACAG 1519  
351 GlyArgSerProSerTyrSerSerThrProSerProLeuMetProAsnLe 367  
1520 GGGCGGTACCTTCCATCAGATTCCACAGCGGTGCGGTATATGCAAACT 1569  
367 uGluAsnPheProTyrSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 384  
1570 GAGGAAGTCTCCATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1619



384 laGlyIleThrAspHisSerHisPheMetProLeuLeuAsnProSerPro 400  
1620 CAGGGATCACTGACCAAGACCTTCCGTCGCCCTTCTGCTTCAATCCCTCCCA 1669  
401 ThrAspAlaThrSerSerValAspThrGlnAlaGlyAsnCysLysProLe 417  
1670 ACGGATGCCACCACTGTGTGGACACCCAGGCTGGCAACTGCAAGCCCT 1719  
417 uGlnLysAspLysLeuProGluAsnLeuLeuSerAspLeuSerLeuGlnS 434  
1720 TCAGAGGACAGGCTCCCTGAGAACCTGCTGGGATCTCAGCTGCGAGA 1769  
434 erLeuThrAlaLeuThrLeuGlnValGluAsnLysSerAsnThrValGln 450  
1770 GCCTCAGCGGCTGACCTCAGCTGAGAGTGGAGAACATCCACACCGTCCAG 1819  
451 GlnLeuLeuSerLysAlaAlaValProGlnLysLysGlyValLysAs 467  
1820 CAGCTGCTGCTCCCAAGGCTGTGTGGCCGAGAGAAAGGTGTCAAGAA 1869  
467 nLeuValSerArgThrProGluGlnHisLysSerGlnHisCysSerProG 484  
1870 CCTGTGTCAGACCCACAGCAGCAGATAAAGCCAGCACTGCAGCCCG 1919  
484 lndLysGlyTyrSerAlaGluProAlaGlyThrProLeuSerGluPro 500  
1920 AAGGAGCGGCTACTCAGCCGAGCCGCGAGCACACCGCTGTCAAGCGG 1969  
501 ProSerSerThrProGlnSerThrHisAlaGluProGlnGluAlaAspTy 517  
1970 CCGAGCAGCAGCCAGCTCCAGCATCGGAGCGGAGCGCGACTA 2019  
517 rLeuSerGlySerGluAspProLeuGluArgSerPheLeuTyrCysAsnG 534  
2020 CCTCAGCGGCTCCGAGACCCAGTGGAGCGAGGCTTCTTACTCAACC 2069  
534 lndAlaArgGlySerProAlaArgValAsnSerAsnSerLysAlaLysPro 550  
2070 AGGCGGTGGAGCCCTGCGAGGGTCAACACAACTCGAGGCCAGGCC 2119  
551 GluSerValSerThrCysSerValThrSerProAspAspMetSerThrLy 567  
2120 GAGTCCGTGTCCACCTGTTCTGTGACCTCTCTGACGACATGTCACCAA 2169  
567 sSerAspAspSerPheGlnSerLeuHisGlySerLeuProLeuAspSerP 584  
2170 AFTCAGGACTCTTCCAGAGCCCTACAGCGAGTCTCCGCTCGACACT 2219  
584 heSerLysPheValAlaGlyLysArgAspCysProArgLeuLeuSer 600  
2220 TCTCCAAAGTTCGTGGCGGTGAGCGGACTGTCCGCGGCTGTCTCAGC 2269  
601 AlaLeuAlaGlnGluAspLeuAlaSerGluIleLeuGlyLeuGlnGluAl 617  
2270 GCCCTGGCAGAGGAGCTGGGCTCCGAGATCTTGGGGCTCGAGGAAGC 2319  
617 alileGlyLysAlaAspLysAlaTrpAlaGluAlaProSerLeuValL 634  
2320 CATCGGTGAGAGGCGGACAAAGCTTGGGCTGAAGCAGCCAGCTGTGTA 2369  
634 ysAspSerSerLysProPheSerLeuGluAsnHisSerAlaCysLeu 650  
2370 AGGACAGCAGCAAGCCACCTTCTCCTGGAGAACCAACAGCCCTGCTG 2419  
651 AspSerValAlaLysSerAlaTrpProArgProGlyGluProGluAlaLe 667  
2420 GACTCTGTGCCAAGGTGGTGGCCCGGCTGGGAGCCAGAGGCCCT 2469  
667 pProAspSerLeuGlnLeuAspLysGlyLysAlaLysAspPheSerP 684  
2470 GCCCGACTCTTGCAGCTGGCAAGCGGCGCAATGCCAAGGACTTCAGCC 2519  
684 roGlyLeuPheGluAspProSerValAlaPheAlaThrProAspProLys 700

2520 CAGGGCTGTTGAAGACCTTCCGTGGCCTTCGCTACGCTGACCCCAAA 2569  
701 DyThrThrGlyProLeuSerPheGlyThrLysProThrLeuGlyValPr 717  
2570 AAGACAACTGTCTCTCTCTTGTGTAACCAAGCCACCTTGGGTTC 2619  
717 oAlaProAspProThrThrAlaAlaPheAspCysPheProAspThrThra 734  
2620 TGTCTCCAGACCCCTACAGCAGCTTTTGACTGTTTCCCGGACACACCG 2669  
734 laAlaSerSerAlaAspSerAlaAsnProPheAlaTrpProGluGluAsn 750  
2670 CTGCCAGCTCAGCGAGCAGCGCAACCCCTTCTGCTGGCCAGAGAAAC 2719  
751 LeuGlyAspAlaCysProArgTrpGlyLeuHisProGlyGluLeuThrLy 767  
2720 CTGGGGATGCTTGTCCAGGTGGGATTTGACCTTGGCGAGCTTACCAA 2769  
767 sGlyLeuGlnGlnGlyLysAlaSerAspGlyLysSerLysGlyAspT 784  
2770 GGCCTGGAGCAGGTGGGAAGCCCTCAGATGCATCAGCAAGGGGACA 2819  
784 hrHisGluAlaSerAlaCysLeuGlyPheGlnGluAspProProGly 800  
2820 CCTCAGGCTTCCGCTGCTGGCTTCCAGGAGGAGGACCCCTCTGG 2869  
801 GluLysValAlaSerLeuProGlyAspPheLysGlnGluValGlyG 817  
2870 GAGAAGGTGGCTCTGTTGCCCGGGACTTCAACGAGGAGGAGGTGGTGG 2919  
817 yValLysGlnGluAlaGlyLysLeuGlnCysProGluValAlaLysA 834  
2920 GGTGAAGAGGAGGAGGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2969  
834 laAspArgTrpLeuGluAspSerArgHisCysCysSerThrAlaAspPhe 850  
2970 CTGACCGGTGGCTGGAGGACAGCGGACTGCTGTTCCACCGCGACTTC 3019  
851 GlyAspLeuProLeuLeuProThrSerArgLysGluAspLeuGluAl 867  
3020 GGGGACTCTCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3069  
867 aGluGlnGluTyrSerSerLeuCysGluLeuLeuGlySerProGluGlnA 884  
3070 TGAGGAGGAGTACTCTCTCTGAGTCTGCTGCTGCTGCTGCTGCTG 3119  
884 rgProGlyMetGlnAspProLeuSerProLysAlaProLeuIleCysThr 900  
3120 GGCTTGGCATGAGGACCGGCTGTCCCAAGGCCCTCATCTATCTGCACC 3169  
901 LysGluGluValGluGluValLeuAspSerLysAlaGlyTrpGlySerPr 917  
3170 AAGGAGGAGTGGAGGAGTGTGACTTCAAGGCCGCTGGGGCTCTCC 3219  
917 oCysHisLeuSerGlyLysSerValIleLeuLeuGlyProThrValGlyT 934  
3220 GTGCCACTCTCAGGGAGTCCCTCATCTGCTGGGCCCTACAGTGGGCA 3269  
934 hrGluSerLysValGlnSerTrpPheGluSerSerLysSerHisMetLys 950  
3270 CCGAGTCAAGTCCAGAGTGTGTTGAGTCTCTCTGCTCAGCATGAG 3319  
951 ProGlyGluGluGlyProAspGlyLysAlaProGlyAspSerThrTh 967  
3320 CAGGTGAGAGGGGCTGATGGGAGGAGGCTTCCAGGGATTCACCAC 3369  
967 rSerAspAlaSerLeuAlaGlnLysProAsnLysProAlaValProGluA 984  
3370 CTGGAGCGCTCTCTGCGCCAGAGCCCAAGCGCTGTGTGCGCGAG 3419  
984 laProIleAlaLysGluProValProArgGlyLysSerLeuArgSer 1000



3420 CGCCATCGCAAGAGAGCTGTGCCACGGGGCAAGAGCTTAACGAGC 3469  
1001 ArgArgValHisArgGlyLeuProGlnIleGlnuSpSerProCysArgAl 1017  
3470 CGTCGGGTGACCGGGGGCTGCCGGGCGAGAGACTCCCAATGAGGAGC 3519  
1017 aProValLeuProLysAspLeuLeuLeuProGlnIleSerCysThrGlyProP 1034  
3520 ACCAGTGTCCCAAGAGCTCTTGTCTCCCTGAATCCTGACAGGGCCCC 3569  
1034 roGlnIleGlnMetGlnIleGlnIleGlnIleProGlnIleArgGlyAlaSerGlu 1050  
3570 CCCAGGGAGAGATGAGAGGGGCTGAGCCCCAGCGCGGGGGCTCGGAA 3619  
1051 GLeuProArgMetCysThrArgSerLeuThrAlaLeuSerGlnProAr 1067  
3620 GGGCTCCCGAGATGTACTGCTCTCTCAGCGGCTTGAGTGAAGCCCCG 3669  
1067 gThrProGlyProProGlyLeuThrThrProAlaProProAspLysL 1084  
3670 CAGCGCCGAGCCCCAGGCTGACCCACCCCTGACCCCGCCAGAACAC 3719  
1084 euGlyGlyLeuGlnArgAlaAlaPheLysSerGlyLysArgValGlyLys 1100  
3720 TGGGGGCAAGCAGAGAGCGCTTCAAGTCGGAACGGGGTGGGGAAG 3769  
1101 ProSerProLysAlaAlaSerSerProSerAsnProAlaAlaLeuProVa 1117  
3770 CCCTCACCCAGGCTGCTCCAGCCCGACAGACCGGGCGCTGCTGG 3819  
1117 AlaSerAspSerSerProMetGlySerLysThrLysGlnThrAspSerP 1134  
3820 GGCTCCGACAGACCCGATGGCTCCAGACCAAGAGAGACAGACACAC 3869  
1134 roSerThrProGlyLysAspGlnArgSerMetIleLeuArgSerArgThr 1150  
3870 CCAGACGCGCTGGCAGACGACGCTCATGATCCTTCGCTCACGCCAC 3919  
1151 LysThrGlnGlnIlePheHisSerLysArgArgArgProSerGlnIleYar 1167  
3920 AAAACCCAGAGATCTCCACCAAGCGGAGGCGCTCTGAGGGCG 3969  
1167 gLeuProAsnCysArgAlaThrLysLysLeuLeuAspAsnSerHisLeuP 1184  
3970 GCTCCCACTGCGTGCACCAAGAGCTCTCGACAAAGCCACTTGC 4019  
1184 roAlaThrPheLysValSerSerSerProGlnIleGlnIleArgValSer 1200  
4020 CCGCCACATTCAGGTCTCCAGACGCCCGCAAGAGGGGAGGTGAGC 4069  
1201 GlnArgAlaArgValProLysProGlyAlaGlySerLysLeuSerAspAr 1217  
4070 CAGCGGCGAGGGTCCCAAGCTGTGTGACGAGCAAGCTCTTGACG 4119  
1217 gProLeuHisAlaLeuLysArgLysSerAlaPheMetAlaProValProT 1234  
4120 GCCCTCCATGCTCCAAAGAGTGGCTTCAAGGGCGCGGCTCCCA 4169  
1234 hLysLysArgAsnLeuValLeuArgHisGlySerSerSerSerSerAsn 1250  
4170 CCAAGAGGAGACCTGTCTTCGAGAGCCGACAGCAGCAGCAGCAAC 4219  
1251 AlaSerAlaMetGlyLeuMetIleArgArgArgGlyLeuArgValProP 1267  
4220 GCCAGTGGCAATGGGGAGATGGAGAGGAGGAGGCTTGGGCTCC 4269  
1267 roProSerSerArgGlyCysLeuLeuProArgLysProSerProProArg 1283  
4270 CACCTCTTCAAGAGAGTCTTCTCCCAAGAAACCAACCCCAAG 4319  
1284 AlaMetAlaSerLeuProGlnSerSerHisProArgArgProProPheL 1300  
4320 GCATGGGAGGCTGCCAAGAGCTCCACCCCGGAGACCCCGCATGCG 4369  
1300 euProGlnAlaArgLeuSerAlaAlaPheGlnGlnIleAlaMetLysThrLy 1316  
4370 TCCCTCAAGCTCCGCTCTCGGGAGGCTTCCAGGGGGCCATGAGACCAA 4419  
1316 sValLeuProProArgLysGlyArgGlyLeuLysLeuGlnIleValG 1333  
4420 GGTGTGTGCACCCCGAAGGGCCGGGGCTCGAAGCTGGAACCATGTGTC 4469  
1333 hLysAlaLeuSerProSerLeuLysLysPheAlaCysLysAlaProGly 1349  
4470 AGAATACACTCGCCCGACCTCAAGAGTTCAGATGAAAGCCCGCAGG 4519  
1350 AlaSerProGlyAsnProLeuSerProSerLeuSerAspLysAspArgG 1366  
4520 GCTCTCTGATATCCTCTGAGCCATCCCTTCGCAAGAGCCGTGG 4569  
1366 yLeuLysGlyAlaGlyLysProValGlyValGlnGlnIleLeuValA 1383  
4570 GCTCAAGGTGTGGGGAGCCCGAGTGGGGGTGGAAGAGGCTGGTAA 4619  
1383 snValGlyThrGlyGlnLysLeuProThrSerGlyAlaAspProLeuCys 1399  
4620 ATGTGGCACCGGGGCAAGAGCTCCCACTTCTGGGCTGATCCCTATGC 4669  
1400 ArgAsnProThrAsnArgSerLeuLysGlyLysLeuMetAsnSerLysLy 1416  
4670 AGAATCCACCAACAGATCTTAAAGCCAACTCATGAACATGAAGAA 4719  
1416 sLeuSerSerThrAspCysPheLysThrGlnAlaPheThrSerProGlnA 1433  
4720 ACTGTCTTACTGACTTTCATAAAGCAGAGCCCTTCACTCCCGGAGG 4769  
1433 lAlaLeuInProGlyGlyThrAlaLeuAlaProLysLysArgSerArgLys 1449  
4770 CCGTCAGAGCTGGGGGAGCTGCGCTGAGCTTAAAGAGAGAGCGGAAA 4819  
1450 GLeuArgAlaGlyAlaHisGlyLeuSerLysGlyProLeuGlnLysArgPr 1466  
4820 GGGCGGCAAGGGGCCATGAGACTCTCCAAAGGCCGCTGAGAGAGCGGCC 4869  
1466 oTyLeuGlyProAlaLeuLeuLeuThrProArgAspArgAlaSerGlyT 1483  
4870 CTATCTTGGCCCGCTCTGCTCTGACTCCCGAGACAGGGGCGAGTGA 4919  
1483 hGlnGlnIleAlaSerGlnuSpAsnSerGlyGlyGlyLysLysProLys 1499  
4920 CACAAGGGGCGAGTGAAGCAACTGTGGGAGGAGCCAGAACCCAAAG 4969  
1500 MetGlnIleLeuGlyProAlaSerGlnProProGlnIleArgProCysG 1516  
4970 ATGAGAGAGCTGGGCTGCTCCAGCCCGGAGGGGAGAGGCTTGGCA 5019  
1516 nProGlnThrArgAlaGlnLysGlnProGlyHisThrAsnLysSerSerT 1533  
5020 GCCCGACAGAGGGCAGAGAACCGAGGCCACCACTACAGCAGGT 5069  
1533 ySerLysArgLysArgLeuThrArgGlyValArgAlaLysAsnThrThrSer 1549  
5070 ATTCACAGCGGAGCGCTCACTCGGGCGGGGCAAGAACACACACTCT 5119  
1550 SerProCysLysGlyArgAlaLysArgArgArgGlnGlnIleValLeuP 1566  
5120 TCACCTGTAAAGGGGCTGCCAAGCGAGCGAGCGAGCGAGGTGGCTCC 5169  
1566 oLeuAspProAlaGlnProGlnIleArgGlyLysThrIleSerSerCysT 1583  
5170 CCGGATCCGACAGAGCTGAAATCCGCTCAAGTACATTTCTCTTGA 5219  
1583 yAsArgLeuArgSerAspSerArgThrProAlaPheSerProPheValArg 1599  
5220 AGGGCTGAGGTGACAGACCGGAGACCCCGCTTTCACACCTCTGA.... 5265

1600 ValGluLysArgAspAlaPheThrThrThrLeuCysThrValValAsnSerPr 1616  
5266 .....  
1616 oGlyAspAlaProLysProHisArgLysProSerSerSerSerAlaSerSerS 1633  
5274 TGGAGGCGCA...AAACCAAGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5320  
1633 erSerSerSerSer.....SerPheSerLeu 1641  
5321 GCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5370  
1642 AspAlaAlaGlyAlaSerLeu..... 1648  
5371 ACGGAGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5420  
1648 ..... 1648  
5421 CTATATATATGTTACATAGATGATATATATGTTGGGACATGCTCGCTTC 5470  
1649 .....AlaThrLeuProGlyGlySerIleLeuGlnProArgPro 1661  
5471 TCCGCTGTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5520  
1662 SerLeu..... 1663  
5521 GCGCTTCCGCGGTAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5570  
1664 .....ProLeuSerSerThrMethHisLeuGlyProValValS 1676  
5571 CGACGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5620  
1676 erLysAlaLeuSerThr 1681  
5621 CCGAGCTCTGTCGCGG 5637  
seq name: gb\_ro:MUSMA  
seq\_documentation\_block: 7222 bp mRNA ROD 07-FEB-1999  
LOCUS MUSMA  
DEFINITION Mouse mRNA for unknown product, complete cds.  
ACCESSION D29801  
VERSION D29801.1 GI:475015  
KEYWORDS ORF,  
SOURCE Mus musculus postnatal-day1 brain cDNA to mRNA, clone\_lib:library  
of Stratagene (La Jolla, CA).  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
AUTHORS Inai,Y.  
TITLE Direct Submission  
JOURNAL Submitted (04-APR-1994) to the DDBJ/EMBL/GenBank databases. Yuji  
Inai, Osaka University Medical School, Molecular Neurobiology,  
Tanabe; 2-2 Yamadaoka, Suita, Osaka 565, Japan (Tel:06-879-3646,  
Fax:06-879-3648)  
2 (bases 1 to 7222)  
REFERENCE  
AUTHORS Inai,Y., Suzuki,Y., Matsui,T., Tohyama,M., Watanaka,A. and Takagi,T.  
TITLE Cloning of a retinoic acid-induced gene, Gt1, in the embryonal  
carcinoma cell line P19: neuron-specific expression in the mouse  
brain  
JOURNAL Brain Res. Mol. Brain Res. 31 (1-2), 1-9 (1995)  
MEDLINE 96078271  
COMMENT Submitted (04-Apr-1994) to DDBJ by:  
yuji Inai  
Department of Molecular Neurobiology, Tanabe  
Osaka University Medical School  
2-2 Yamadaoka, Suita  
Osaka 565  
Japan  
Phone: 06-879-3646  
Fax: 06-879-3648  
Location/Qualifiers

Source  
1. 7222  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone\_lib="library of Stratagene (La Jolla, CA)"  
/dev\_stage="postnatal-day1"  
/tissue\_type="brain"  
309.5831  
/codon\_start=1  
/product="Unknown"  
/protein\_id="BAA06184.1"  
/db\_xref="GI:475016"  
/translation="MOSPRRCGFGHKQOYVPTSOETSRLENYRQPGAGLSCDRQR  
LLADYISQPVYGVGGTGPSTVATAADKVRHSGSLQGRPAFPYSVQDSSYP  
GRYSEGLQWGGPP  
RTHSLYPP  
VQGGCGGSAHSYKSTCAQPSAOPHRMSANLAPQVQNLHAYQPRGLYEQQQAL  
QGRHOTOELHYNLAQHYQGGQGYGCTPTAVTPEQYQYQYQYQYQYQYQYQYQYQY  
RSPYSSTPLMPLNENFYSOQPLSTGAFPTGTDHSHFMPLLNPSTDAASVDP  
QAGNCKPLKEKLPNLLSYLSQSLTALTQVNIQVNIQVNIQVNIQVNIQVNIQVNIQV  
LVSTPEQHKSCPEGSGYSAEPAGTPLSEPPSPSPSPSPSPSPSPSPSPSPSPSPSP  
ERSFLYCSQARGSPARVNSKAPSPVSCVSTSPDNDSTKSDSFQSLHSTLPLPS  
FSKTVAGERCPLLLSALAQEDLASEILQEAIVEKADKAWAEASSLPKDNKPPF  
SLENHGACLDIVAKTSWQGPETPLPEPLQDKGSKDFSPGLFEDFSVAFATDP  
KTSSTPLSGTGLLTATPPTAAQDFDPTPTASSVDCGANFPANPEENLGDACPR  
WGLHPGLTKLEOGAKADGVGKADAHASACMGFOEDHAIKPAALSGDFKQOEA  
EGVKEVGLLQCPVAKADONLESHHSCSSTDFGLPLPPGKEDLEAEERYSS  
LOELGSPQRPLODPLSPKAPLMCTKEAEALDTRAGVWSPCHLSEPAVLLGSP  
VGAQSKVSWFESSLHMKPGEGEREMERAPGSGTSQSLAPRPNKPAVPPEPIAK  
BEVPRGLSRVRHGLPEAEADSPCPALPKDLLEPCTGPPQGGAGAGPGRG  
LSEGLPRMCTRSITALSPPOTPGPGLTTPTPDKLGKQRAAFSGRRVGPSPKA  
ASSPSNPAALPVASDSSPMGSKTEKDPSPMGKQORSWLRRTKPPQVFAKRLRP  
SERIPDCRATKPLPANNHLPATKVSQSGPQKQSGRQVVKPQKPGNKLSDRLHT  
LKRKSAMFAPVPAKESLILSNNGSGDGEREAPSPGLLRMASPORAPRSGGE  
RVAGAPPTPRSPALPERRPGSPAGAEGLGGMGOMLPAASGADPLCRNPASRLK  
KLNSKLLSNAADCPKAEAFNSPTPLSLGTARAPKRSKRGTTGLPSKPLEKRP  
CGQPLLLAHDRASTOGGCDNNGSGGKKPKTELGAPSPQPPQPPQPPQPPQPPQPPQ  
PCQASTYSIKRRLSRGRGTAAHSPCKGKATRRRQOQVLPDPAEPRLKLYSSC  
KRLRADSTPAFSPFVVEKRDYITCTVNSPGEFAPHKWSPSSSAASSTSSSL  
EPAGASLFTTFGSLQQRSLPSTHMLGPVYSKALSTCLVCCICQNPANFKDLG  
DLCGPYVPEHCKPKKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEK  
ITPTVALGRSLRPPDPADKOGPLRTSARGLSRRLOSCYCDGCGGGEVAAQADKS  
RHEKSEAPTEFGGTQSEHWHEACAVTSGVYLAVAGLGLQEAQMKVAVDMVRG"  
7202..7207  
polyA\_signal  
polyA\_signal 7208..7213  
polyA\_site 7222  
BASE COUNT 1612 a 2335 c 2027 g 1248 t  
ORIGIN  
alignment\_scores:  
Quality: 7504.50 Length: 1762  
Ratio: 4.621 Gaps: 15  
Percent Similarity: 92.168 Percent Identity: 81.725  
alignment\_block:  
US-09-508-821B-6 x MUSMA ..  
Align seg 1/1 to: MUSMA from: 1 to: 7222  
1 MetGlnSerPheArgGluArgCysGlyPheHisGlyLysGlnGlnAsnTy 17  
309 ATGCAGTCTCTTTTCGAGAAAGGTGTGTTTCCATGCGCAACAGCAAACTA 358  
17 xGlnGlnThrSerGlnGlnThrSerArgLeuGluAsnTyArgGlnProS 34  
359 CCCACAGACTCCCGAGGACATCGCGCTCGAGACTACAGCAGCCGG 408  
34 erGlnAlaGlyLeuSerCysAspArgGlnArgLeuAlaLysAspTyr 50  
409 CTCAGGCTGGGCTAAGCTGTGATCGGCGAGCGGCTGCTGGCGAAGACTAC 458  
51 TyrAsnProGlnProTyrProSerTyrGluGlyGlyAlaGlyThrProSe 67

```

||||| 459 TACAGCCCTCAACCTATACAGGCTATGAGGATGGACCTGACACCTTC 508
67 rglYThraAlaAlaValAlaAlaAspLysTyrHisArgGlySerLysA 84
509 TGGACACGGTGGCCACAGACCTGCACACAAAGTACACACGAGGACAAAT 558
84 lAdeuProthrgInglnglyLeuGlnGlyArgProAlaPheProGlyTyr 100
559 CC.....CTGCAGGGAGGACGACCTTCCCTCCAGCTAT 590
101 GLyAlaGlnAspSerSerProGlyTyrProGlyTyrAlaGlnGluSe 117
591 ...GTCAAGACAGACCCCTACCCAGGAGCTACTCTGGAGAGAGG 637
117 rLeuGlnAlaTrpGlyAlaLapProGlnProProGlnProGlnProL 134
638 TCTTCAGACCTGGGGGGGCCACAGACCGCTCCCTCCAGCAGACCTC 687
134 euproAlaGlyValAlaLysTyrAspGluAsnLeuMetLysLysThrAla 150
688 TGCCGTGGGGAGTGAAGCAAGTGAAGAACCTGATGAAGAAAGACTT 737
151 Val...ProProSerArgGlnTyrAlaGlnGlnGlyValAlaProPh 166
738 GTGCCCTCCCAACAGCAGTACCTGAGCAGGGCCCGCAGCTCCCTT 787
166 eArgThrHisSerLeuHisValGlnGlnProProProProGlnGlnProL 183
788 CCGGACTCACAGCTGCATGTC...CCACACACACAGCTCCAGCAGCCCC 834
183 euAlaTyrProLysLeuGlnArgGlnLysLeuGlnAsnAspIleAlaSer 199
835 TGGCTTACCCCAACTCCAAAGGAGAAACACAGAACGACCTGCCTCC 884
200 ProLeuProPheProGlnGlyThrHisPheProGlnHisSerGlnSerPh 216
885 CCTCTGCCCTTCCCGCAGGAGCAGCTTCCCGCAGCATCCCGACCTT 934
216 eProThrSerSerThrTyrSerSerSerValGlnGlyGlyGlnGlyA 233
935 CCTACCTCTCCACTTATGCCCCAACATGAGGTGGTGGGCGAGGGG 984
233 lAnHisSerTyrLysSerCysThrAlaProThrAlaGlnProHisAspArg 249
985 CCCACTCTCCACAGAGCTGCACAGCACACTGCTGCCAGCTCATATAGG 1034
250 ProLeuThrAlaSerSerSerLeuAlaProGlyGlnArgValGlnAsnLe 266
1035 CGATGAGTGCATATGCAAGCTGGCTCCAGGGGCAACGGGCTCCAGAACT 1084
266 uHisAlaTyrGlnSerGlyArgLeuSerTyrAspGlnGlnGlnGlnG 283
1085 TCAAGCTTACAGACCTGCGCCCTTGGCTAC..... 1115
283 lngGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 299
1116 .....GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1154
300 GlnGlnThrLeuHisTyrGlnAsnLeuAlaLysTyrGlnHisTyrGly 316
1155 CAGGAACAACCTCCACTACAGAACTCGCAGATACCAACAGATAGAGCA 1204
316 nGlnGlyGlnGlyTyrCysGlnProAspAlaAlaValAlaThrProGln 333
1205 GCAAGGCAAGGGCTACTGCTCCAGGAGACAGCTGTCAAGGACCCAGAGG 1254
333 lntTyrGlnThrPheSerProSerSerSerHisSerProAlaArgSer 349
1255 AGTATACAGAGCTTCAAGCCGAGCTCCAGGACACTCCCTCCAGAGGCT 1304
350 ValGlyArgSerProSerTyrSerSerThrProSerProLeuSerProAs 366
|||||
1305 GTGGGTCCCTCCCTCCATTAAGCTCCACCCCGCTCACACAGTACGCCAA 1354
366 nLeuGlnAsnPheProTyrSerGlnGlnProLeuSerThrGlyAlaPhe 383
1355 TGTGGAGACTTCCCTATAGCCACAGACCGCTTATATGGGGCTTCC 1404
383 rAlaGlyTlPheThrAspHisSerHisPheMetProLeuAsnProSer 399
1405 CCAGAGGAGATCACAGACACAGACCTTATAGCCCTGTAAACCCCTCC 1454
400 ProThrAspAlaThrSerSerValAspThrGlnAlaGlyAsnCysLysPr 416
1455 CCAGCATGCTGCCACCTGTGTGAACCCCGAGGCGGAGATGAGAGCC 1504
416 OleuGlnLysAspLysLeuProGlnLysLeuLeuSerAspLeuSerLeuG 433
1505 CCTGCAAAAGGAGAAAGCTGCTGACAACTCTCTCCGAGTGCAGCTGC 1554
433 lnsertLeuThrAlaLeuThrLeuGlnValGlnLysnLleSerSerThrVal 449
1555 AGAGCTCACAGCGCTCAGCTCCAGGTGGAACATCTCCACACCTG 1604
450 GlnGlnLysLeuLeuSerLysAlaAlaValProGlnLysLysGlyValLys 466
1605 CAGACCTCTTCTGTCGAAGCTACATGCACAGACAGAAAGGGGTCAA 1654
466 sAsnLeuValSerArgThrProGlnGlnHisLysSerGlnHisCysSerP 483
1655 GAACTCTGTGTCCAGACTCCAGACAGACACAAAGAGCAGCTGTAGCC 1704
483 rGlnGlySerGlyTyrSerAlaGlnProAlaGlnGlyThrProLeuSerGln 499
1705 CTAGAGGACAGCGGCTACTGAGTGCAGCCAGCGGGCACACCGCTGTCAG 1754
500 ProProSerSerThrProGlnInsertThrHisAlaGlnProGlnGlnAlaAs 516
1755 CCGGACAGCAGCAGCGCCAGCTCCACCTGCTGACACAGAGACCTGA 1804
516 pTyrLeuSerGlySerGlnAspProLeuGlnLysSerPheLeuTyrLysA 533
1805 CTACAGTGTGCTGTGAGACCCCTGAAACGAGACCTCTCTCTCTCA 1854
533 snGlnAlaArgLysProAlaArgValAsnSerAsnSerLysAlaLys 549
1855 GCCAGGCGCGGGGAGCGCCCGCAGGGTCAACACAACTCAAGGCTTAAG 1904
550 ProGlnSerValSerThrCysSerValHisSerProAspAspMetSerTh 566
1905 CCCAGCTGTGTCCACCTGTGTGTGACCTCAGCTGATGATGATGATCCAC 1954
566 rLysSerAspAspSerPheGlnInsertLeuHisGlySerLeuProLeuAsp 583
1955 CAAGTCGACGACTCTTCCAGAGCTGAGACAGTACAGTGCCTGGAGCA 2004
583 ePheSerLysPheValAlaGlyGlnAlaGlyArgCysProArgLeuLeu 599
2005 GCTTCTCCAAATTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2054
600 SerAlaLeuAlaGlnGlnLysPheLeuAlaSerGlnLysLeuGlyLeuGln 616
2055 AGTCCCTGGCACAAGGAAGATCTGGGCTCCGAGATCTGGAGTGCAGGA 2104
616 uAlaTlGlyGlnLysAlaAspLysAlaTrpAlaGlnAlaProSerLeuY 633
2105 AGCCATTTGTTGAGAGGCTGACAAAGCTGGGCTGAGGCTTCCACCTGC 2154
633 llyAspSerSerLysProProPheSerLeuGlnLysnHisSerAlaCys 649
2155 CCAAGGACATAGCAGACGACCTTCTACATGAGGAACATAGGCGCTGC 2204
650 LeuAspSerValAlaLysSerAlaTrpProArgProGlnProGlnAl 666
2205 CTGAGACACGTGAGCAAAAGACTTCAATGTCACAGCCAGGGAGGACAGAG 2254

```

966 rThrSerAspAlaSerLeuAlaGlnIlyProAsnIlySerProAlaValProG 983  
1313155 CACCTTCACAGGCTCTTGCCCGGCAAGGCTTAACAAGCTTGCTTACCTG 3304  
983 luAlaProIleAlaIlyLysGluProValProArgGlyLysSerLeuArg 999  
3323025 AGGGGCCCATTCGTAAAGAAAGAGCTGTGCCACGGGTGAAGAGCTTACGG 3254  
10080 SerArgGlyValHisArgGlyLeuProGlnAlaGluAspSerProCysAr 1016  
3333225 ACCCGGAGTACACCGGGGCTGCTTGAAGGGAAGATCTTCCTGCGAG 3304  
1016 gAlaProValLeuProIlyLysAspLeuLeuLeuProGlnSerCysThrGlyP 1033  
3333305 GTGCGCAGCACTTCCCAAGAGACCTTGTCTCCAGAGTCTTGCACAGGC 3354  
101033 roProGlnGlyGlnMetGluGlyValadlyAlaProGlyArgGlyAlaSer 1049  
3333355 CCCACAGGAGCAGCCGAGGGGCTGGAGCCCAAGCCCGGGGGCTGTCA 3404  
101050 GluGlyLeuProArgMetCysThrArgSerLeuThrAlaLeuSerGluPr 1066  
3343405 GAAGGGCTCCCCAGAAATGTGACCCGCTCGCTTACAGTCTCGAGTGAGCC 3454  
101066 oArgThrProGlyProGlyLeuThrThrThrProAlaProProAspL 1083  
3343455 CAGAGACTCTGGACCTCCGGGCTGACTACACCCCAAGCCCTCCGACA 3504  
101083 ysLeuGlyGlyLysGlnArgAlaAlaPheLysSerGlyLysArgValGly 1099  
3353505 AACTGGAGTAGTAGCAGCAGCTGCTTCAAGTGTGCAAGCGGGCTCGGA 3554  
11100 LysProSerProLysAlaAlaSerSerProSerAsnProAlaAlaLeuPr 1116  
33555 AAACCATCACCGAGCTGCATCCAGCCCAAGCAACCTGCTGCCCTGCC 3604  
1111116 oValAlaSerAspSerSerProMetGlySerIlyThrLysGluThrAspS 1133  
33605 TGTTGCTCTCAGACAGCAGCCCATCTCCAGTCCCAAGCAAGCAAGCAGCAGCT 3654  
1111133 exProSerThrProGlyLysAspGlnArgSerMetIleLeuArgSerArg 1149  
33655 CTCCCAGCATGCTCGCAAGGACCAAGCGCTCCATGCTCTCGGCTCGC 3704  
1111150 ThrLysThrGlnGluIlePheHisSerLysArgArgArgProSerGluGl 1166  
33705 ACCAAACCCAGCAGAGCTTCCATGTCCAACCGGCGCGCCCTCCGAGAG 3754  
111166 YArgLeuProAsnCysArgAlaThrLysLysLeuLeuAspAsnSerHisL 1183  
3373755 CGGATCCAGACTCGCTGCGCCCAAGAGCTCCCTGCTTACACCACTT 3804  
1111183 euProAlaThrPheLysValSerSerSerProGlnLysGluGlyArgVal 1199  
33805 TACCCATCTGGCTTCAAGTCTCCAGTGGCCGCCAAGAGAAAGTCCGGATG 3854  
1200 SerGlnArgAlaArgValProLysProGlyAlaGlySerLysLeuSerAs 1216  
33855 AGCCAGCGGTCAAGTACCCCAAGCCTGGTACAGAAACAAGCTTCAGA 3904  
1216 pArgProLeuHisAlaLeuLysArgIlySerSerAlaPheMetAlaProValP 1233  
33905 TCGGCTCTCCACAGGCTCAGAGGAAGTCCGGCTTTCATGGCACCTGTC 3954  
12123 roThrLysLysArgAsnLeuValLeuArgHisGlySerSerSerSer 1249  
33955 CGCCCAAAACAGGCTCATCTCTGAGA. ....AGCAAC 3989  
1250 AsnAlaSerAlaMetGly. GluMetGlyArgArgArgGlyLeuArgValP 1266  
3990 AATGAAGT. ....GGGGGAGACAGCGGAGGAGAGGCGCTGAGAGTTC 4033  
1266 roProProSerSerArgGlyCysLeuLeuProArgLysProSerProPro 1282

```

4034 CCTGGCCCTCTACGAGAGATGGCTACACCCAGAGGCGAGACCGAGG 4083
1283 ArgAlaMetAlaSerLeuProGlnSerSerHisProArgArgProPh 1299
4084 GTAGTG.....GGAGCCACCCCGACCC..CCACCCCT 4114
1299 eLeuProGlnAla.....ArgLeuSerAlaAlaPheGln 1311
4115 GGAGCCACCTGTCATGCATGGGGTGTCTACACAGCATCTCCCTCCCA 4164
1311 LysAlaMetLysThrLysValLeuProProArgLysGlyArgLysLys 1327
4165 GTGCTGTGAGGACCAAGTGTCTGCCACCGGAAAGGCGCCGCTCCAG 4214
1328 LeuGluAlaIleValGlnLysIleThrSerProSerLeuLysLysPhe 1344
4215 CTGGAGGCGCATGTACAGAAATACCTACCTGGCTCCAGAAAGACTGCG 4264
1344 acLysLysAlaProGlyAlaSerProGlyLysProLeuSerProSerLeu 1361
4265 ATGCAGAGGTGGCAGGGGCCCTCTGGAGACCCCGAGGCCAGCCCTGC 4314
1361 eAspLysAspArgLysLeuLysGlyAlaGlyLysProValGlyAla 1377
4315 CTGAG.....AGAGTCCAGGGGGGCGATGCCAGCCGGGGCG 4349
1378 GlnGlnGlyLeuValAlaGlnValGlyThrGlyLysLeuPro...ThrSe 1393
4350 GAAGAGGGTTTA.....GGAGGAAATGGGCAATGCTACACAGCATCTTC 4393
1393 rGlyAlaAspProLeuLysArgAspProThrAspArgSerLeuLysGly 1410
4394 CGGAGCTGACCATGTGTGAGAAACCCAGCAGCAGTCTCTAAAGTGA 4443
1410 LysLeuMetAspSerLysLysLeuSerSer...ThrAspCysPheLysThr 1425
4444 AACTCTTGAACAGTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4493
1426 GlnAlaPheThrSerProGlnAlaLeuGlnProGlyGlyThrAlaLeu 1442
4494 GAGGCTTCATGTCCCGAGAGACCTCCCATCTAGGAGACTCCCGAGAC 4533
1442 aProLysLysArgSerArgLysGlyArgAlaGlyAlaHisGlyLeuSer 1459
4544 ACCGAGAGAAAGAGCGCGGAAAGGAGAGAGCTGGAGCTGGAGCTCCA 4593
1459 LysGlyProLeuGlnLysArgProLysLysLysProAlaLeuLeuThr 1475
4594 AAGGTCGCTGGAGAGAGCGCGCTGCGGAGCACTGCTGCTGCTGCTGCT 4643
1476 ProArgAspArgLysSerGlyThrGlnGlyAlaSerGlnLysPheSer 1492
4644 CCCACGACAGGCGCAGCAGCAGCAGGAGGAGGAGGAGGAGCACTCCAG 4693
1492 LysGlyGlyLysLysProLysMetGlnLysLysLysProAlaSerGlnP 1509
4694 TGGAGAGGCGCAAGAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4743
1509 rOProGlnGlyArgProCysGlnProGlnThrArgAlaGlnLysGlnPro 1525
4744 CCCCGAAGGCGCGCTGCGACGCCCAAGACAGAGCGCGAGAGAGCGCG 4793
1526 GlyHisThrAsnLysSerSerLysLysArgLysArgLeuThrArgL 1542
4794 GGCCAGGCGAGCTACAGAGCTATCCAGCGGAGGCGCTCCAGCGAGG 4843
1542 ValGlnAlaAsnThrThrSerSerProCysLysGlyValGlnAlaLysArg 1559
4844 CCGGGGAGAGAGCGCGAGAGCTTACCTCTGTAAGAGAGAGAGAGAGAG 4893
1559 rGArgGlnGlnGlnValLeuProLeuAspProAlaGlnProGlnLysArg 1575

```

```

4894 GAAGGACGACAGAGTACTGCCCCCTGAGTCTCGACAGAGCTGAATCCGA 4943
1576 LeuLysThrLysSerSerCysLysArgLeuArgSerAspArgThrPro 1592
4944 CTCAAATACATTCCTCTGAGAGAGGCTGGGAGAGAGAGAGAGAGAGAG 4993
1592 AlaPheSerProPheValArgValGlnLysArgAspAlaPheThrThr 1609
4994 AGCTTCTGCGCCCTTGGCGGGGTGGAGAGAGAGAGAGAGAGAGAGAG 5043
1609 LysThrValValAlaAsnSerProGlyAspAlaProLysProHisArgLys 1625
5044 TATCACTGTCTGCAATCCCGGGGAGAGAGAGAGAGAGAGAGAGAGAG 5093
1626 ProSerSerSerAlaSerSerSerSerSerSerSerSerPheSerLeu 1642
5094 CCAATCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5140
1642 PAlaAlaGlyAlaSerLeuAlaThrLeuProGlyLysSerLysLeuGlnP 1659
5141 ACCAGCTGGGGCTTCTGACACATTCCTGAGAGAGAGAGAGAGAGAGAG 5190
1659 rOArgProSerLeuProLeuSerSerThrMetHisLeuGlyProValVal 1675
5191 AGAGGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5240
1676 SerLysAlaLeuSerThrSerCysLeuValCysCysLeuCysGlnAsnPr 1692
5241 TCCAGAGCCCTAAGTACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5290
1692 AlaAsnPheLysAspLeuGlnLysLysLysLysLysLysLysLysLys 1709
5291 GCCAATTTCAGAGACCTTGGGAGACTCTGAGGCTCTATACCTGAGAC 5340
1709 LysCysLeuProLysLysLysLysLysLysLysLysLysLysLysLys 1725
5341 ACTGCTCCGCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5390
1726 GlyThrCysGlnGlnAlaSerLeuProLeuGlnLysArgThrLeuLys 1742
5391 GGCACCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5440
1742 GlnCysAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1753
5441 GGAATGTGGCGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5474
seq_name: gb-prim.HSM801433
seq documentation block:
LOCUS HSM801433 2660 bp mRNA 18-FEB-2000
DEFINITION Homo sapiens mRNA, cDNA DKFZp344A139 (from clone DKFZp344A139).
ACCESSION AL133649
VERSION AL133649.1 GI:6599191
KEYWORDS
SOURCE
ORGANISM
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1. (bases 1 to 2660)
AUTHORS Poustka, A., Klein, M., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
TITLE Direct Submission
JOURNAL Submitted (15-DEC-1999) MIPs, Am Klopferstr 18a, D-82152
Martinsried, GERMANY
COMMENT
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ), Email: s.wiemann@dkfz-heidelberg.de,
sequenced by DKFZ (German Cancer Research Center,
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
This clone (DKFZp344A139) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de Further
information about the clone and the sequencing project is available
at http://www.mips.biochem.mpg.de/proj/cDNA/.

```







1266 AATCTGCCACCAAGATACCTCTGCAAAAGCCAGGTTGGCAG...TACAA 1312  
332 sGlnProAspAlaValArgThrProGluGlnTyrThrGlnThrPheS 339  
1313 CCAGCCTGAGGTCTCTAGGTCTCCCATGCTCCAGTTCACCAAGATTCA 1362  
339 erProSerSerHisSerProAlaArgSerValGlyArgSerProSer 355  
1363 GCCTATTCTAACCTCTCTCAGTCTCTCTGCTGCTTCTCAGTCTCCAGC 1412  
356 TyrSerSerThrProSerProLeuMetProAsnLeuGluAsnPheProTy 372  
1413 TGAGCTCTACACCTTCTCTCTCATGCTCAGAGTGGTGGAGATCTCCAGT 1462  
372 rSerGlnGlnProLeuSerThrGlyAlaPheProAlaGlyLleThrAspH 389  
1463 TGGGCAA.....GGCAGTGGCCCTCAGTCTCCAGAAACC 1497  
389 is...SerHisPheMetProLeuAsnProSerProThrAspAlaThr 404  
1498 GATTTTACAGTATGCCCCACCTCAGTCCCACTCCATCA...ATGATG 1544  
405 SerSerValAspThrGlnAlaGlyAsnCysLysProLeuGlnLysAspLy 421  
1545 CCAGTCTTAACCTCTCATGCTGCAGGATTCAAAGGATTTGGATTAGAAG 1594  
421 sLeuProGluAsnLeuLeuSerAspLeuSerLeuGlnSerLeuThrAlaL 438  
1595 GGTCCGGAAGGCGGTGACTGATCTCTGGTGTGAGTAGTTGAGTGCTT 1644  
438 eUrThrLeuGlnValGluAsnLleSerAsnThrValGlnGlnLeuLeu 454  
1645 TGAGTACTCAGTGGCCATCTCTCCATTAATCTGTCACGACATGTTCATT 1694  
455 SerLysAlaValProGlnLysLysGlyValLysAsnLeuValSerAr 471  
1695 TCTGATGCTTGACACCTCAGAGAGAGGCTCTAAGAGGCCCTCATCA... 1742  
471 gThrProGluGlnHisLysSerGlnHisCys...SerProGluGlySerG 487  
1743 .....TCATTAAGAAAGCAGATAGCTGTACAACTCAGAGGCTCT 1785  
487 LyTyrSerAlaGluProAlaGlyThrProLeuSerGluProProSerSer 503  
1786 CACAGCCTGAGACACAACTAAGTCCCTATG..... 1817  
504 ThrProGlnSerThrHisAlaGluProGlnGluAlaAspTyrLeuSerG 520  
1818 .....GCTGAGTCTATTGGATGGAGGTTCTCTAGTAG 1849  
520 ySerGluAspProLeuGluArgSerPheLeuTyrCysAsnGlnAlaArg 537  
1850 TTCAGAGATCAAGGTGAGAG..... 1871  
537 LySerProAlaArgValAsnSerAsnSerLysAlaLysProGluSerVal 553  
1871 ..... 1871  
554 SerThrCysSerValThrSerProAspAspMetSerThrLysSerAspAs 570  
1872 .....GTGAGGCAACTGAGTGGCCAGACGACTAGCTCTGACAC 1909  
570 pSerPheGlnSerLeuHisGlySerLeuProLeuAspSerPheSerLysp 587  
1910 CACCTATAAGTGT.....GGCGCTTCAGAGAAAGTGGCTCTCCACAA 1953  
587 heValAlaGlyLysArgCysProArgLeuLeuLeuSerAlaLeuAla 603  
1954 CACAAGGTCTCAGAAAGGAGGCCCTTAGGCTCAGTACCACTGCTGCACT 2003  
604 GlnGluAspLeuAlaSerGlu..... 610  
2004 AGGAGCAGAGTGCCTCTCCAGGTGCTAAGGACACATCACTGTCATCTGA 2053

611 .....IleLeuGlyLeuGlnG 616  
2054 GGGGAACACAAAATCAATGAGAAGACAGTGTGGGTGATTGTCTCCGAG 2103  
616 luAlaIleGlyGluLysAlaAspLysAlaTyrAlaGluAlaProSerLeu 632  
2104 AAGCATGACAGTGGGTGAGAAATCTGTGGGACAGATAAAGGCTCC 2153  
633 VallysAsp.....SerLysProProPheSerLeuGluAs 645  
2154 CAGAGATGATCTCTGTCAGTCCAGAGACCACCT..... 2189  
645 rHisSerAlaCysLeuAspSerValAlaLysSerAlaTyrProArgProG 662  
2190 AGCAATAGTGGGTAAAGAAATCAGCCACACATCACTTCCACAGCCAG 2238  
662 LyGluProGluAlaLeuProAspSerLeuGlnLeuAspLysGlyAsn 678  
2239 AFCTCTCCAGGAGA.....GGGAGCAAGGAACAAGAATGGTATAT 2282  
678 AlAlaLysAspPheSerProGlyLeuPheGluAspProSerValAlaPheAl 695  
2283 AATAGCTTAACCAATGAGAGGGAATGGCCGAGTAGCCACTCCGCTCC 2332  
695 aThr.....ProAspProLysLysThrThrGlyP 705  
2333 AGTAGGCCCAAGTTTCCAGGAGGAGTGCAGCTAGCAAGTCTCTCGAA 2382  
705 rOleuSerPheGlyThrLysProThrLeuGlyValProAlaProAspPro 721  
2383 GTTCCGCTATATTCAAGAGAGTTTGGTGCAGCTACCA...CGA 2429  
722 ThrThrAlaAlaPheAspCysPheProAspThrThrAlaAlaSerSerAl 738  
2430 AATGTCAGTGTATCTCATGATCTCT.....TCAGG 2461  
738 aspSerAlaAsnProPheAlaTyrProGluGluAsnLeuGly..... 752  
2462 ACAAGAAAGGGGATTTGGCAGCCATGGAGAGCGAAGAGGTAGAAATG 2511  
753 .....AspAlaCysProArgTyrProGlyLeuHis 761  
2512 AGAGTCTCCAGTCTTCTACAGGAGTGTCTCAGGGGTACCCACCACAT 2561  
762 ProGlyGlu.....LeuThrLy 767  
2562 CCTGACAGAGGATCTCTAGAGCCGACAGGAACATCAAGGGATGGCTAG 2611  
767 sGlyLeuGlnGlnGlyLysAlaSerAspGlyLleSerLysGlyAspT 784  
2612 TGSCCTGGAA.....GGAATGCAAGGCTTAACATCTTAGTCAGTCAAA 2655  
784 hrHisGluAlaSerAlaCysLeuGlyPheGlnGluGluAspProProGly 800  
2656 CCAATGAATTAGTAGC..... 2672  
801 GluLysValAlaSerLeuProGlyAspPheLysGlnGluValGlyG 817  
2673 .....AGGGGCT 2680  
817 yValLysGluGluAlaGlyLysLeuGlnCysProGluValAlaLysA 834  
2681 TCTGAACAAGAGTATGGTATCCCTGTGGAAATCCC..... 2717  
834 laAspArgTyrLeuGluAspSerArgHisCysCysSerThrAla..... 848  
2718 .....CAGTGGGACATGGGAAGGAAGTCAAGCAGCAGCTCTCTGNA 2762  
849 .....AspPheGlyAspLeuProLeuProThrSerAr 861  
2763 ATGAACAGATCAATTTCTGTGATCTCCATT.....CCGAG 2800



861 glysluaspueugluhlaugluuyluyrserleucysgluend 878  
 ||| : : : : :  
 2801 GAAGTTGAGTAGAAGCTCCA..... 2822  
 878 euGlyserProgluInarProgluMetInasProleuSerProlys 894  
 ||| : : : : :  
 2823 ....TCATCAGCCCATGAGCCTGSA....GGCTCCCTTTCTGAAAG 2861  
 895 AlaProleuileCysThrlysgluInuValgluValleu..... 908  
 : : : : :  
 2862 AGSTCAGTAATCTGTATATATCCCAAGACATGTCAGAGGACC 2911  
 909 .....Asperlysalaglytrp 915  
 : : : : :  
 2912 AGGGCTCACTCATGAGACACATGGTACTGATGCCAATATGGAGGA 2961  
 915 lyserProCysHisleuser...GlyluserValleleuendlyPro 930  
 : : : : :  
 2962 ATGAAAGCTCAACCCCAAGTTTAAGTCAGTCATCTTCTGTCGG 3011  
 931 ThrValglyThrlyserlyValInsertrp..... 941  
 : : : : :  
 3012 TTGCTCTCATGAAACAAATGAAATCCAGGGTGACAGATAAAGA 3061  
 942 .....PhegluserSerleuserHisMetlyserProglygluInuGlyP 956  
 : : : : :  
 3062 GGAAGACTTTGACAAATCCAAATCTAAGCTAGTTTAAACAAATCTG 3111  
 956 roASP.....Glylu 959  
 : : : : :  
 3112 GAGACCACTGCCATCCTACAGCATCAACATGAGACTTACGTGCAAT 3161  
 960 ArgAlaProgly.....AsperthrThrSerAspAlaSerleuAl 973  
 : : : : :  
 3162 GCAAGTCTGAGAGCAGACGCCCATGATTCATTCAGACATGAGCCCA 3211  
 973 agluInyProasnlyPro.....AlaValProgluAlaProleuAla 988  
 : : : : :  
 3212 AGATPAGAGGTCCACACCATGCGGGGCTCCT...GATAGAGTGGTA 3258  
 988 yslsgluProValProArglylyserleuArgserArgValHis 1004  
 : : : : :  
 3259 GCCGGAGACTATG...AGGGCTGGGTCC....TCCTTCAGTACAT 3299  
 1005 ArgglyleuProgluAlaInasSerPro.....CysArgAlaProVa 1019  
 : : : : :  
 3300 GACTTTCAGAAAAATGAGATGTCTCAGGACAGACAGAGGCCAGG 3349  
 1019 lleuProlyAspLeuLeuLeuProgluSerCysThrlyProProglu 1036  
 : : : : :  
 3350 GGGAGACCTCATCATGATGACCCCAT...ATGACCTTTTCAGAGAGG 3396  
 1036 lygluMetgluInuValaglyAlaProglyArgglyAlaSerlyleu 1052  
 : : : : :  
 3397 CCAATGAGAGTCTTACATGCTCTTTTCTCCCACTCAGAAAGCTG 3446  
 1053 ProArgMetCysThrArgserleuThrAlaLeuSerluProArgThp 1069  
 : : : : :  
 3447 GCCTCTGCTTACACACA.....CAAGCAC 3469  
 1069 rogly.....ProProgly 1073  
 : : : : :  
 3470 CAGGCTCATGCTTATGGGACCTATACGGTTTGAATTCAGAGTCC 3519  
 1074 LeuThrThrThrProAlaProProAspLyseu.....G 1086  
 : : : : :  
 3520 ATTATAGAGACAGATGTACCAACAGCAAGAGAGATTAAGACTGG 3569  
 1086 lylysgluArgAlaAlaPheLyserlylyArg..... 1097  
 : : : : :  
 3570 GCGACAG.....TTCTGCTCAGGGAGTATGCTGTCGACGCA 3610  
 1098 ...ValglyLysProser.....ProlysalAlaSerProser.A 1111

3611 TAGGACGAGAGGCCACAGCAGAGCCCAAGCACA...GCATTTCTG 3657  
 : : : : :  
 1111 snProAlaAlaLeuProValAlaSerAspSer..... 1122  
 : : : : :  
 3658 ACAGATTAAGGAGCCCTGAAAAATGACAAAGAGGAAATGATGATGC 3707  
 1123 ...ProMetglySer.....LysThrlygluThrAspSer... 1133  
 : : : : :  
 3708 CCACAGTAGACGTATCATGACCCACACACTGAGAGACCTGGGCGCTG 3757  
 1134 ...ProSerThrProglylyAspAlaProInasSerMetleuArg...S 1148  
 : : : : :  
 3758 TCTCATGTCTAGATGAGTCTGCTGCCCTCCCAAGCATGATTAAGCACA 3807  
 1148 eArgThrlyThrInuInuIlePheHis.....SerlyArgArgArg 1162  
 : : : : :  
 3808 GCTCTCAGAAAGTTACAAAGATGCTGTTGGAGATCTTTCGAGACTCT 3857  
 1163 ProsergluInuArgleuProasnCysArgAlaThrlyslsleuLeuAs 1179  
 : : : : :  
 3858 CCAGCCAAAGCAGTGTCTCCAGAAATGTATATCAAAAGGTATGG 3907  
 1179 pasSerHisleuProAlaAlaThrPheLysValSerSerProgluInu 1196  
 : : : : :  
 3908 GCCACCCATGACCCAGTGAAGATGAGACTTACCCAGTCTGACATCAT 3957  
 1196 luGlyArgValSerInuArgAlaArgValProlysProglyAlaGlySer 1212  
 : : : : :  
 3958 CCAAACCTAGTACGTACATGTTACAGCTT.....CTGTCAGAGAGAT 4001  
 1213 LysleuserAspArgProleuHisAlaLeuLysArglySerAlaPhe 1229  
 : : : : :  
 4002 CATCTCTCAAAATCCTTATATCATCGAAGAGGGGTGCTTTTAT 4051  
 1229 tAlaProValProThrlyLysArgAsnLeuValleuArgHisLysers 1246  
 : : : : :  
 4052 CTCCCTATTCAGTAAAGACGTACACAAGTGA... 4091  
 1246 erSerSerAsnAlaSerAlaMetgluInuMetlyArgArgly 1262  
 : : : : :  
 4092 ....AACAGTAAAGCTGATGATTAAGGGGCCCTC..... 4121  
 1263 LeuArgValProProProserSerArglyLysleuLeuProArgly 1279  
 : : : : :  
 4122 .....CTCCACCTCAAAAGA 4138  
 1279 oSerProArgAlaMetAlaSerleuProgluInasSerHisProArg 1296  
 : : : : :  
 4139 AGGTGCCGATTAACATACAAATCTTACAGGCATCTTCTCAGATCAG 4188  
 1296 rgyProPheLeuProgluAlaArgleuser..... 1306  
 : : : : :  
 4189 ATATCAAGTCTATTCCTTAAGAGAGATTCCTCAAGACCTTCCAAACCA 4238  
 1307 .....AlaAlaPheInuValMetlyThrly 1316  
 : : : : :  
 4239 GACATATGAACTGCTGCTGTATACCTGCAAGTCTGCTTAACACCA 4288  
 1316 sValleuProProArglyslArglylyleuLysleuInuAlaIleVal 1333  
 : : : : :  
 4289 AATATGCCCCCAGAGAGAGGAGGATTAATTTGAAGCATATAGTTC 4338  
 1333 InuylleThrSerProserleuLysPheAlaCysLysAlaProgly 1349  
 : : : : :  
 4339 AGAAGATCATCTTCCAAATATAGAGAGAGCATGTGCAAAAGCTCT 4388  
 1350 AlaSerProglyAsnProleuSerProserleuserAspLysAspArg 1366  
 : : : : :  
 4389 GAGGCTGGGGAGACAGTCA...CTGATGATAT 4423  
 1366 yLeuLygluAlaGlylyserProValgly..... 1376

[illegible]

```

seq_name: gb_pr:HS597B2

seq_documentation_block:
LOCUS       HS597B2          DNA                  PRI      12-DEC-1999
DEFINITION  Human DNA sequence from clone RP4-597B2 on chromosome 22 Contains
            an exon of the TCF20 gene for transcription factor 20
            (transcription factor 20) complete sequence.
            (transcription factor 20) complete sequence.

```

AL031346  
AL031346.8 GI:4760290  
HTG: ABI: KIAA0292: TCF20.

REFERENCE: 1 (bases 1 to 43761)

REQUEST: On May 7, 1999 this sequence version replaced gi:582316. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

<http://www.sanger.ac.uk/HGP/Chr22>

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Embl, EMBL; Swi, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information



```

repeat_region 21378..22070
/notes="figger3b repeat: matches 2..741 of consensus"
repeat_region 22071..22382
/notes="Alusx repeat: matches 1..309 of consensus"
repeat_region 22383..22393
/notes="figger3b repeat: matches 741..751 of consensus"
repeat_region 22394..22673
/notes="Alusx repeat: matches 1..282 of consensus"
repeat_region 22674..22840
/notes="figger3b repeat: matches 751..928 of consensus"
repeat_region 22841..23138
/notes="Alusx repeat: matches 1..297 of consensus"
repeat_region 23139..23585
/notes="figger3b repeat: matches 928..1231 of consensus"
repeat_region 23738..24015
/notes="Alusx repeat: matches 1..313 of consensus"
repeat_region 24067..24116

```

## alignment\_scores:

Quality: 993.50 Length: 2067  
 Ratio: 1.027 Gaps: 96  
 Percent Similarity: 46.783 Percent Identity: 25.109

## alignment\_block:

US-09-508-821B-6 x HS597B2/rev ..

Align seg 1/1 to reverse of: HS597B2 from: 1 to: 43761

```

1 MetGlnSerPheArgGluArgCysGlyPheHisGlyLysGlnGlnAsnTyr 17
|||||
6451 ATGCAGTCCCTTCGGGAGCAAGAGGTTACACGGAACCAACCAAGCTA 6402

17 rGlnGlnThrSerGlnGlnThrSerArgLeuGlnAsnTyrArgGlnPro 34
|||||
6401 CCCACAGAGGTTACAGGCTCATCCGCTAGAAGAGTTC...AGCCCTC 6355

34 exGlnAlaGlyLeuSerCysAspArgGlnArgLeuAlaLysAspTyr 50
|||||
6354 GTCAGGCC..... 6347

51 TyrAsnProGlnProTyrProSerTyrGlnGlyAlaGlyThrProSe 67
|||||
6346 .....CAGATGTTCCAGATTTT...GGAGGTACAGGTGCGCATAG 6309

67 rGlyThrAlaAlaValAlaAlaAspLysTyrHisArgGlySerLysA 84
|||||
6308 TGCCAGCAGTGGCAGT...GGCAGTGGTGGTGACGACGAGCAGCAG 6262

84 lAlaLeuProThrGlnGlnGlyLeuGln.....GlyArgProAlaPhePro 98
|||||
6261 CTGCTCCGCGAGCGTGGCTACGAGACCTCTGCCATCAAGTTACCAG 6212

99 GlyTyrGlyValGlnAspSerSerProTyrProGlyArgTyrAlaGly 115
|||||
6211 GGTTCAGGAAGAGGCTGGAGATTTTAC.....TACATGGCAGCA 6168

115 uGluSerLeuGlnAlaTrpGlyAlaProGlnProProProGlnProG 132
|||||
6167 CAAGAAGCCCGTACTACAGAACCCACAGCCT.....CCTC 6130

132 lProLeuProAlaGly...ValAlaLysTyrAspGluAsnLeuMetLys 147
|||||
6129 AGCAGAGCTCTTCGGCCCTGTCAGACCTATGGA..... 6095

148 LysThrAlaValProPro.....SerArgLntTyrAlaGly 159
|||||
6094 .....CCCCCAGGGAGGACAGCTTTGGCAATCAGTATGGAG 6057

159 uGlnGly.....AlaGln..... 163
|||||
6056 TGAGGTCTATGTGGCCAGTTCACAGCAGCACTCTGGCCTTGGCGTG 6007

164 .....ValProPheArgThrHisSerLeu 171

```

```

6066 TGTCACATTATCAGCAGGATTACACTGGGCTTTCTCTCCAGGGAGTGCT 5957
|||||
172 HisValGlnGlnProProProGlnGlnProLeuAlaTyrProLysLe 188
|||||
5956 CAGTACCAACAGCAGGCTTCAGCCAGCAGCAGCAGCAGCAGCAGCAGCA 5907
|||||
188 uGlnArgGlnLysLeuGlnAsnAspLeuAlaSerProLeuProPhePro 205
|||||
5906 GTTGAGAGCAACAGCTTTACCAG.....TCCCATCAGCCCTGCCAC 5866
|||||
205 lGlyThrHisPhePro.....GlnHisSerGlnSerPhePro 217
|||||
5865 AGGCCACTGGCCAAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 5816
|||||
218 ThrSerSerThrTyrSerSerValGlnGly.....Gln 229
|||||
5815 CGGCCCTCACTCTGCCATCCTCTGCTGCTGTACCAAGTTAAGAGTGG 5766
|||||
229 yGlyGlnGlyAlaHisSerTyrLysSerCysThrAlaProThrAlaGlnP 246
|||||
5765 TCAGTTTGGCCAAACACTATCAGTCTTCTCTCTCTCTCTCTCTCTCT 5716
|||||
246 rHisAspArgProLeuThrAlaSerSerSerLeuAlaProGlyGln... 261
|||||
5715 CCTTCCCTTCCACACAGCGTTTACCGAGTCT.....GGACAGGC 5675
|||||
261 ..... 261

5674 TATGATGCCAGTTACATATGTAATCTGGATCTCAGTATGAAGGACACA 5625
|||||
262 .....A 262

5624 TGTGGTCTTAATGCACAGGCTTATGGAACACATCCAAATTACAGCTATC 5575
|||||
262 rGValGlnAsnLeuHisAlaTyrGlnSerGlyArgLeuSerTyrAspGln 278
|||||
5574 AGCCTCACTATGAAGATTTGACAGGCAAGATTCACACAGGGACC 5525
|||||
278 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 295
|||||
5524 CAACAGGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 5486
|||||
295 rArgHisHisAlaGlnGlnThrLeuHisTyrGlnAsnLeuAla...LysT 311
|||||
5485 .....CACCTCTCTCAGCATGTGATGCGATATACAGCTGCCACCAAGC 5440
|||||
311 yGlnHisTyrGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 327
|||||
5439 TGCCCTCGAAGCCAGTGGGCAG...TACACAGCAGCTGAGGTCTCT 5393
|||||
328 ValArgThrProGlnGlnTyrTyrGlnThrPheSerProSerSerSerH 344
|||||
5392 GTGAGGTCCCATCAGTCTTCCAGAGACTCAGCCCATTTCTTAACCC 5343
|||||
344 sSerProAlaArgSerValGlyArgSerSerTyrSerSerTyrProSer 361
|||||
5342 TTCTCCAGCTGCTCTGTGGTTCAGTCTCCAAAGCTGATGTTCTACCCAT 5293
|||||
361 eProLeuMetProAsnLeuGlnAsnPheProTyrSerGlnGlnProLeu 377
|||||
5292 CTCCTCTCATGAGCTGGGAGATCTCCAGTGTGGGCAA..... 5252
|||||
378 SerThrGlyAlaPheProAlaGlyLysThrAspHis...SerHisPhe 393
|||||
5251 .....GGCAGTGTCCATGTTCCAGAGAACAGAAATTTTACAGTTAT 5208
|||||
393 lProLeuAsnProSerProThrAspAlaThrSerSerValAspThrG 410
|||||
5207 GCTCAACTCAGTCCACCAACCCATCA...ATGATGCCAGTCTTAATTC 5161
|||||
410 lAlaGlyAsnCysLysProLeuGlnLysAspLysLeuProGluAsnLeu 426
|||||

```

5160 ATGCTGAGGCTTCAGAGGTTGGACCTAGAGAGGCTACCAAGGCA 5111  
 427 LeuSerAspLeuSerLeuGlnSerLeuThrAlaLeuThrLeuGlnValG1 443  
 5110 CTGACAGATCTCGGTGAGAGTTGAGTGTCTGTGAGTACCAAGTGGC 5061  
 443 uAsnIleSerAsnThrValGlnGlnLeuLeuSerLysAlaIleValP 460  
 5060 CAATCTTCTTACACTGTCAGACACATGTTACTTCTGAGGCCCTGACTC 5011  
 460 roGlnLysLysGlyValLysAsnLeuValSerArgThrProGlnHis 476  
 5010 CTCAGAGAGAGACCTCCAG.....AGGCCCTCATCTTCCAG 4973  
 477 LysSerGlnHisCys...SerProGlnLysSerGlyTyrSerAlaGluP 492  
 4972 AAACAGATAGCTGCACAAATCTGAGAGCTCTCCACACCTGAAGACA 4923  
 492 oAlaGlyThrProLeuSerGlu.....ProProSer 503  
 4922 GCTGAAGTCCCTATGCGAGATCATTAAGAGAGCTGCTCCACAGATT 4873  
 503 erThrProGln.....SerThrAlaGluPro 512  
 4872 CAGAGATCAAGCGAGAGAGTCCGCAACTAGTGCACAGACACAGC 4823  
 513 GlnGlnAlaAspTyrLeuSerGlySerGlnAspProLeuGlnArgSerP 529  
 4822 TCTGACACACCTACAGAGGAGCTTGTGAGAAAGCTGCTCTCTC... 4775  
 529 eleuTyrCysAsnGlnAlaArgLysSerProAlaArgValAsnSerAs 546  
 4774 ....CCGGCACAAAGTGTCTGAGATGAACCCCGACCTCAATGCTACTC 4729  
 546 erLysAlaLysProGlnSerValSerThrCysSerValThrSerProsp 562  
 4728 CTGCCGCAAGAGAGAGGCC.....ACCTCCACCA... 4700  
 563 AspSerSerThrLysSerAspSerPheGlnSerLeuHisGlySerLe 579  
 4700 ..... 4700  
 579 uProLeuAspSerPheSerLysPheValAlaGlyLysArgSerProA 596  
 4699 .....GGCGTAAAGGACATGCCA. 4682  
 596 rGleuLeuSerSerAlaLeuAlaGlnLysLeuAlaSerGlnLeu 612  
 4681 .....TTGTCAATCCGAGGAGAACCCAAAGGTTATGAGAAAGCTTT 4640  
 613 GlyLeu.....GlnGlnAlaIleGlyLysAlaLysAlaTr 626  
 4639 GGGGTGATGTCTCCGGGAGGACATGAGGTGGGTAAAGAACCTGG 4590  
 626 PALGlnAlaProSerLeuValLysAsp.....SerSerLysP 639  
 4589 TGGACAGATTAAGGCTCCCAAGAGATATCTGACAGCTCAAGAGC 4540  
 639 roProPheSerLeuGlnAsnHisSerAlaCysLeuAspSerValAlaLys 655  
 4539 CACCT.....AGCAATGCTGGGGCAAGGAACCACTATCATCA 4499  
 656 SerAlaTyrProArgPro.....GlyGluProG1 665  
 4498 CTTCCTCCAGCCAGACCTCCAGAGAGAGAGGAGCAAGAGAAACAA 4449  
 665 uAlaLeuProAspSerLeuGlnLeuAspLysGlyLysAlaLysAsp 682  
 4448 TGGCGATTAACAACCTCAACATATGAGAAAGAGAGAGAGTGGCC 4399  
 682 heser.....ProGlyLeuPheGluAspProSerValAlaPheAla 695  
 4398 ACTCTGAGAGGGGCCCTGCTTT.....ACGAGC 4370

696 ThrProAspProLysLysThrThrGlyProLeuSerPheGlyThrLysP 712  
 4369 AGACAGACCTACCAATCTCTGAGAGCTCGCTATATGATCAAGAG 4320  
 712 cThrLeuGlyValProAlaProAspProThrThrAlaIleAspCysP 729  
 4319 TAGTTCGGGTGACCCCTGCCA...CGAATGACATGGCTTCTCCAGT 4273  
 729 heProAspThrThrAlaIleSerSerAlaAspSerAlaAspProPheAla 745  
 4272 ATCTT.....ACAGGCAAGAGAGAGAT.....TTCACT 4241  
 746 TrpProGlnGlnAsnLeuGlyAspAlaCysProArgTyrPglyLeuHisP 762  
 4240 GGCCATGGGAGAACAAAGGCT.....AGAAATGAAATATCC 4203  
 762 oGlyGlnLeuThrLysGlyLeuGlnGlnGlyLysLysAlaSerAspLys 779  
 4202 AAGCTCTCGCAGAGAGTCTTCAGGTTACACACACCTACAGAGA 4153  
 779 leSerLysGlyAspThrHisGlnAlaSerAlaCysLeuGlyPheGlnGlu 795  
 4152 GATATTCTGAGGTACTCAAGACATCAGGGAGTGGCTGTAAGCTAGAA 4103  
 796 GluAspProProGlyGlyLysValAlaSerLeuProGlyAspPheLysG1 812  
 4102 GGAACCAAGGCCCAATGCTTGTGTGTAACCAATGATTAAGTCTAG 4053  
 812 nGlnGlnValGlyGlyValLysGlnGlnAlaGlyLysLeuLeuIncSP 829  
 4052 CAGG.....GGCTCTGAAACAAAGATGGGTGCTCATTAAGAAATC 4009  
 829 roGluValAlaLysAlaAspArgTyrPheGlnLysAspSerArgHisCys 845  
 4008 CC.....CACTGGGGCCCTGGGAGAGGAAATCAAGC 3977  
 846 SerThrAla.....AspPheGlyAspLeuProLeu 856  
 3976 AGCAGACTCTGAAATGAAGACATCAATTTGACTATACCAAT... 3929  
 856 uProProHiserArgLysGlnAspLeuGlnAlaGlnGlnGlnLysP 873  
 3928 .....CCCAAGAGTTGAAATAGACCTCGAG..... 3902  
 873 erLeuGlnLeuLeuGlySerProGlnGlnArgProGlyLysGlnAsp 889  
 3901 .....TCAACAGACATAGACCTGG.....GGT 3878  
 890 ProLeuSerProLysAlaProLeuLecysThrLysGlnGlnValGlnG1 906  
 3877 TCCTCTCTGAAAGAGATCAGTATCTGTATATTTCTCCACTAAGACA 3828  
 906 uValLeuAspSerLysAlaGlyTyrPglySerProCysHisLeuSer... 921  
 3827 GATTGTC..AGGACCCAGGGCTCCTACTCTCGGACACATAGAGTCCG 3781  
 922 .....GlyGlnSer 924  
 3780 ACAACGAAATTGGAGAGATGACCTCAATCAACTTAAGTCAAGTCTG 3731  
 925 ValIleLeuLeuGlyProThrValAlaGlyThrLysSerVal... 938  
 3730 GTCAATCTCTCGGTGTGTGTGTCTGAGAAAGACCTCAATATCCCA 3681  
 939 .....GlnSerTyrPheGlnLys 945  
 3680 GAGCGGAGATTAAGAGAGAGACTTTGAGACAGTCAATATCAACTCA 3631  
 945 erLeuSerHisMetLysProGlyGlu..... 953  
 3630 GTTTCACACAGAAATCTGAGAGACCTGCTCACTCTTAAGATCAAG 3581



2066 TGCCATGACATCCACAACTCAGAGAACCTGTAGTAAACCAAGGAGAC 2017  
 1452 la...gylalahlisglyleuserlysylproleuglulysargprotyr 1467  
 2016 CAGATGTTCCCTGGT ..... 2000  
 1468 leuglyproalaleuleuethrproargaspargalaserglythr 1484  
 1999 ...GGAAAGACACCTTAATCTTCAGACCAAAAGATGATACCTCCAGT 1953  
 1484 ngl.....Alaserglaspasenserglyglyglyglylys... 1497  
 1952 GGGCATATTGGCCCTGAGCAACCCCAAGCTGAGAGAGAGAGACG 1903  
 1498 .....Prolysmetglugluleuglyproalaser 1508  
 1902 ATACAGTACGATTTCCAGCAAGAGGTTCCCTCCAGAGGATAT 1853  
 1509 propro.....gluglyargprotyrproglulthrargal 1521  
 1852 TTCCATCAGAAAGAGAGAGAGACCATTTGATGATGATTAAGCA 1803  
 1521 agllysgln..... 1524  
 1802 AAAGAAAGACAGACGACCCGCTCCACCCCTCAGCCCAAGATAC 1753  
 1525 ..Proglylthrasnlyrserserlysarglyshargleuthr 1540  
 1752 CAGAGGTTCTGCAGATGAGAGACCAAGCCAAAGAGAGAGAGG 1703  
 1541 Argglyargalalysasnthrthrserserproctylslyglyargalaly 1557  
 1702 AGGAGAGAGAGAGAG.....CTGGGGCCAGCCAGAGAA 1668  
 1557 sargargarglnglnglnvalleuproleuasprralagluuprogli 1574  
 1667 GCGAAAGACCAAGACAGCTGCTCCATGTCAGAACCCAGAGACCTGGA 1618  
 1574 leargleustyrlyrlesersercylsargyleuargseraspserarg 1590  
 1617 TCMACTAAATATCCACCCAG...CCACTGGATTAATATCTATCCAG 1571  
 1591 ThrProalapheserProphavalargvalglulysargspalapheth 1607  
 1570 AACAGCTTTTACCTTACATCCATGATTAATATGATGAGACTGG 1521  
 1607 rthliecysthrvalalasnserproglyasp..... 1618  
 1520 AGCCGTTGTACATCATGTCAGAGAGAGAGAACCAATATG 1471  
 1619 .....Alaprolysprohialarglyseraseraseralaser 1633  
 1470 TGAAGGAGAGAGAGGTCAGAGCTCAGACCTCCACCTGAGAGACT 1421  
 1634 SeraseraseraserPheaserleuaspalaladialaserleualath 1650  
 1420 GAAAGCAG..... 1412  
 1650 rleuproglyglyserileleuglinproargproserleuproleusers 1667  
 1411 .....GCCCTCCGGCTCGT 1396  
 1667 erthmethislenuglyprovalalaserlysalaaleuserthsercys 1683  
 1395 CTTTATGCTGAGAGGACCTGTTGTACAGAGCTCTGATTAAGGAGC 1346  
 1684 leuvalcyscysleucysglinasnproalasnphelyaspheuglyas 1700  
 1345 CTGGTTGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1296  
 1700 pleucysglyprotyrtyrproglulhis.....Cysleuprolysl 1714  
 1295 CTTCTTTGACCTTTTATCCCAAGATTAAGAGCAGCCATCTCCGAGA 1246

1714 yslsypso.....lysleuylsleuylsvalarg 1723  
 1245 ATCCACCTCTTAAGAGGCGCACAGAAATGACAGAGCAAGTAAAGTACG 1196  
 1724 progluglymthrcysgluglulnlaserleuproleuunargthrleul 1740  
 1195 CAC.....AAAGTCTCTTAATGAGCTCCCAAGAGAGAGAGCT 1158  
 1740 ys.glyproglucysalaalalalathralaglylserpropro 1754  
 1157 GGAG 1113  
 seq\_name: gp\_sts:AF021115  
 seq\_documentation\_block:  
 LOCUS AF021115 595 bp DNA STS 03-APR-1998  
 DEFINITION Homo sapiens trinucleotide repeat ctg-23, sequence tagged site.  
 ACCESSION AF021115  
 VERSION AF021115.1 GI:3015588  
 KEYWORDS STS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 595)  
 Philbert,R.A., Horelli-Kultunen,N., Robb,A.S., Lee,Y.H.,  
 Long,R.T., Damschroder-Williams,P., Martin,B.M., Brennan,M.B.,  
 Palotie,A. and Gimsu,E.I.  
 The characterization and sequence analysis of thirty CTG-repeat  
 containing genomic cosmid clones  
 Eur. J. Hum. Genet. 6 (1), 89-94 (1998)  
 98454304  
 JOURNAL MEDLINE  
 2 (bases 1 to 595)  
 Philbert,R.A., Horelli-Kultunen,N., Robb,A.S., Lee,Y.H.,  
 Long,R.T., Damschroder-Williams,P., Martin,B.M., Brennan,M.B.,  
 Palotie,A. and Gimsu,E.I.  
 Direct Submission  
 Submitted (28-AUG-1997) Clinical Neuroscience Branch, National  
 Institutes of Mental Health, 49 Convent Drive Rm B1E16, Bethesda,  
 MD 20892, USA  
 FEATURES  
 source location/qualifiers  
 1..595  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="17"  
 /map="17q11.2"  
 /note="primers to amplify repeat: forward:  
 ACAGCACGATGCTCCGAGCCCAT; reverse:  
 GCTGATCTTGGGAGGTTTGGT"  
 BASE COUNT 90 a 140 c 230 g 120 t 15 others  
 ORIGIN  
 alignment\_scores:  
 quality: 873.50 length: 199  
 ratio: 4.671 gaps: 6  
 Percent Similarity: 93.970 Percent Identity: 88.945  
 alignment\_block:  
 US-09-508-821b-6 x AF021115/rev ..  
 Align seq 1/1 to reverse of: AF021115 from: 1 to: 595  
 178 PROPROGLINGINPROLEU..AlATyrProLySLeuGlnArgGlnLysLeu 194  
 533 CCGGCCAGACAGCCCTGNGCATAACCCAGACCTCCAAAGAGAGAGCTGC 544  
 134 lnsnaspilalaserProleupropheproglulnlythrhisphero 210  
 543 AGAAGCATTCCTCCCTGCTGCTCCCTCCAGGAGGATACCATTTCT 494  
 211 GlnHisSerGlnSerPheProThrSerThrThrThrSerSerValG 227

[illegible]





860 SerArgLysGluAspLeuGluAlaGluGluLysrSerSerLeuCysG1 876  
1793 CCCAGAAAGTTGAATAGAGCTCAG ..... 1819  
876 uLeuLeuGlySerProGluGlnArgProGlyMetGlnAspProLeuSerP 893  
1820 ..... 1853  
893 rOlySalProLeuLysCysThrLysGluValGluGluValGluLeuAsp 909  
1854 AAGAAGATCAGTGCATGTCGATGATTTCTCCACTAAGACAGATGTC ... 1900  
910 SerLysAlaGlyTrpGlySerProCysHisLeuSer ..... 921  
1901 AGGGACCCAGGGGCTCACTCACTGGGACACATGAGTCCGACACAGAAAT 1950  
922 ..... GlyCysValIleLeuL 928  
1951 TGGGAGGAATGACCTCTCAATCACTTAAAGTCAGTGGTCTCTTCTTC 2000  
928 euGlyProThrValGlyThrGluSerLysVal ..... 938  
2001 CTGGTGGTTGGTTCATGGAACCAAGCTGAATCCAGAGCGGGCAG 2050  
939 ..... GlnSerTrpPheGluSerLeuSerHi 948  
2051 AATAAGAGGAGACTTTGAACAGCTGAATCTCAAGCTAGTTTCAACAA 2100  
948 sMetLysProGlyGlu ..... GluGlyP 956  
2101 CAAGAATCTGGAGACCACTGCCATCTCTAGCTCAAGCATGAGTCT 2150  
956 roAspGlyGluAlaProGlyAspSerThrThrSerAspAlaSer ... 971  
2151 ACCGGGCAATGCGACCTCGGACGACCAACCATGATTCCTTTCAGAC 2200  
972 LeuAlaGlnLysProAsnLysProAla ..... ValProGluAl 984  
2201 TATGGCCCGCAAGACGACGACCCACGCCAATCGCGGGTCCCT...GG 2247  
984 aProLeuAlaLysGluProValProArgGlyLysSerLeu ..... 998  
2248 CAGAGTTGGTGGTGGGAGGGGATG...AGGGGTGGTCCCTTCTCAAT 2294  
999 ..... ArgSerArg ... 1001  
2295 ATCAGCACTTTCAGAAAATTTGAAATGCTCTCTGGCGGAGCAGAGGC 2344  
1002 ..... ArgValHisArgGlyLeuProGlu... 1009  
2345 CCAGGGGAGACCTCATCATGAATCCACATGACCTTTTCAGAGAG 2394  
1010 AlaGluAspSerProCysArgAlaProValLeuProLysAspLeuLeu 1026  
2395 GGCTAACCGGAGTCTTTACACACTCCCTTTTCTCCCACTCAGAAACC 2444  
1026 eu ..... 1026  
2445 TGGCTCTCTATCATGCAATCTCGGGCTCATGTTGGGACCT 2494  
1026 ..... 1026  
2495 AACGCAGTTTGAATCTCAGCTGCATTATAAGACAGATGTAACA 2544  
1027 ..... ProGluSerCysThrGlyProGlnGlyGlnMetGluGlyAlaG 1042  
2545 GCACCCAGAGAGTATAAGACTGGAGCAGCGGTCTGCTCAGGGGTAA 2594  
1042 lValAlaProGlyArgGlyAlaSerGluGluLeuProArgMetCysThrArg 1058  
2595 TTGCTGCAGCAGCAGCAGGAGGAGGG...CCAGGAAGAGTCCAAAG 2641

1059 SerLeuThrAlaLeuSerGluProArgThrPro ..... 1069  
2642 CAGCAGCAGTTCTTTCAGACAGTACGGAGCCCTCTGAAAAATGACAAAGA 2691  
1070 ..... GlyProPro...GlyLeuThrThrThrProAlaPro 1080  
2692 TGGTATGATGATGCGCCACCACTGGGACTTACCACGCCAGTGGCC 2741  
1081 ProAspLysLeuGlyGlyLysGlnArgAlaAlaPheLysSerGlyLysAr 1097  
2742 ..... AGGAG 2746  
1097 gValGlyLysProSerProLysAlaAlaSerSerProSerAsnProAla 1114  
2747 GCTGGGC ..... 2753  
1114 lAlaLeuProValAlaSerAspSerSerProMetGlySerLysThrLysGlu 1130  
2754 ..... GCTGCCTAATCTAGTGTGTGCTGCTTACACAGGCCA 2792  
1131 ThrAspSerProSerThrProGlyLysAspGlnArgSerMetIleLeuAr 1147  
2793 TGGAAAT ..... TAAAGCATGGCTCCAG 2815  
1147 gSerArgThrLysThrGlnLysPheHisSerLysArgArgArgProS 1164  
2816 AAGTTACAGAATCTCTTGGGATCTTCTCGCAAACTTCTCCAGCCAA 2865  
1164 erGluGlyArgLeuProAsnCysArgAlaThrLys ..... 1175  
2866 AAGCAGCGTCTCTCCAGGAATGCCAGTCAAAAAGGTATGGCGGCC 2915  
1176 ...LysLeuLeuAspAsnSerHisLeuProAlaThrPheLysValSerSe 1191  
2916 ATGAGACTGATGGACATGGACTAGCTGAG ..... GCTAC 2949  
1191 rSerProGlnLysGluGlyArgValSerGlnArgAlaArgValProLysP 1208  
2950 ACAGTCATCAAACTGCTGTAGTGTATGCTGAGACTT ..... C 2987  
1208 roGlyAlaGlySerLysLeuSerAspArgProLeuHisAlaLeuLysArg 1224  
2988 CAGCCAGGAGGATCATCTTCTCAAAACCCCTTAATCATGAGGAGCGT 3037  
1225 LysSerAlaPheMetAlaProValProThrLysLysArgAsnLeuVal 1241  
3038 GTTGGTCTTTTATCTCTCCATCCAGTAAGACAGACAGTCAACAGATGT 3087  
1241 uArgHisGlySerSerSer ..... 1248  
3088 AAGAAGACAGTACACTGAAGATAAAGTGCCTCTTCACTCATCAAAAG 3137  
1249 ..... SerAsnAlaSer 1252  
3138 AAGCGCTGATTAAGCATTCAATCTTATGCCCACTTCTCAGACTCAG 3187  
1253 AlaMetGlyGluMetGlyArgArgGlyLeuArg...ValProProPr 1268  
3188 GATATCAATCTATCCCTAAGAGATTTCTCCAAAGACCTTCCAAAGTCC 3237  
1268 sSerSerArgGlyCysLeuLeuProArgLysProSerProArgAlaM 1285  
3238 AGATAGTAGAAACTGCCCTGTGTATCCCTCAAGCCCT ..... 3277  
1285 eAlaSerLeuProGlnSerSerHisProArgArgProProPheLeuPro 1301  
3277 ..... 3277  
1302 GlnAlaArgLeuSerAlaAlaPheGlnGlyAlaMetLysThrLysVal 1318  
3278 ..... GCTAAGACCAAAATACT 3294  
1318 uProProArgLysGlyArgGlyLeuLysLeuGluAlaIleValGlnLysI 1335

```

|||||
3295 GCGCCGACGAGAGAGCGGGATTGAATGGAAGCTATAGTTCAGAGAA 3344
1335 lcthrserProserLeuLysPheAlaCysLysAlaProglLysaser 1351
|||||
3345 TTAAATCCCCAAATATAGAGAGAGCGCATCTCGAAGACGTGGAGGCT 3394
1352 ProglYasnProLysSerProSerLeuSerLysAspArgLysLeu 1368
|||||
3395 GGGGAGAGACAGGTTACG.....CTTGATGATATACTGCTTTGAA 3435
1368 sgLYala.....GlyGlySerProValGlyAluGlyLeu 1382
|||||
3436 GAGTGTCTCTCTGAGAGTGGAGCT...GTTCCTGTTGAGAGTGGAGA 3482
1382 AlaenValGlyThrGlyLysLeuProThrSerGlyAlaAspProLeu 1398
|||||
3483 TAGAGAGAGAGAAAGGTGAGGTG.....GCTTGAGACTTA 3517
1399 CysArgAsnProThrAsnArgSerLeu.....LysGlyLysLeuMetAs 1413
|||||
3518 GTC...AGTCAGACAAACCAAGAGATTGACATAGAGAAACCTCTCCAG 3564
1413 nserLys.....LysLeuSerSerThrAspCysPheLysThrGlu 1426
|||||
3565 GTCTTCAGAGAGAGTGGCGGTGGACCGTGAATGCAAGTGAAGACAGAA 3614
1427 ....AlaPheThrSerProgluAlaLeuGlnProglYgLy..... 1438
|||||
3615 CACATGCAAGAAACAGTTACTGCGGAAAGAGAACCCCTGGTGCCATGACA 3664
1439 ThrAlaLeuAlaProLysLysArgSerArgLysGlyArgAla...GlyAl 1454
|||||
3665 TCCACAACTCCACAGAGCTGTGTAACCAAGAGAGACCAAGTGGTTTC 3714
1454 atHisGlyLeuSerLysGlyProLeuGlnLysArgProGlyLeuGlyProA 1471
|||||
3715 CCGGGT.....GGAAACAG 3728
1471 lAlaLeuLeuThrProArgAspAlaGalaSerGlyThrGlnGly..... 1485
|||||
3729 CACCTTTAATCTTCCAGACTCAAGAAATGACTCCAGTGGGCAATATG 3778
1486 AlaSerGluAspAsnSerGlyGlyGlyLysLys..... 1497
|||||
3779 GCGCCTGAGCAACCCCAAGGCTGAGAGAGAGAGAGAGAGAGAGAGAG 3828
1498 .....ProLysMetGlnGlnLeuGlyProAlaSerGlnProPro.... 1510
|||||
3829 GATTTCACCGAAGAGAGGGTTTCCCTCAAGAGGATATTTCCCATCAG 3878
1511 .....GlnGlyArgProCysGlnProGlnThrArgAlaGlnLysGln 1524
|||||
3879 GAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3928
1525 .....ProGlyH1 1527
|||||
3929 CAGCAGCCAGCGCCTCCACCCCTCAGCCCAAGAGATACAGAGAGAGTTC 3978
1527 sthrAsnLysSerSerLysArgLysArgLysArgLysArgLysArgLys 1544
|||||
3979 TGCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 4028
1544 lAlysAsnThrThrSerSerProCysLysGlyArgAlaLysArgLysArg 1560
|||||
4029 GGAAG.....CTGGGGCCAGCCGAGAGAGAGAGAGAGAGAGAGAG 4063
1561 GlnGlnGlnValLeuProLeuAspProAlaGlnProGlnLysArgLeu 1577
|||||
4064 AATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 4113
1577 styLieserSerCysLysArgLeuArgSerAspSerArgLysProAlaP 1594
|||||

```

```

4114 AATGCCACCCAG...CCACTGATATAAAGTATGTCAGAGAGAGAGTCTT 4160
1594 lAserProPheValArgValGlnLysArgAspAlaPheThrThrLeu 1610
|||||
4161 TTACCCCTTACATCCATGATGATATAGTGAATCTTGAGCGCTTTCT 4210
1611 ThrValAlaAsnSerProglYasp.....AlaP 1620
|||||
4211 ACATATCATCATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4260
1520 oLysProHisArgLysProSerSerSerAlaSerSerSerSerSer 1637
|||||
4261 GAAAGGTGAGAGAGTACATACCTCCACCTACAGAGAGAGAGAGAGAG 4309
1637 erSerPheSerLeuAspAlaValGlyAlaSerLeuAlaThrLeuProglY 1653
4309 ..... 4309
1654 GlySerThrLeuGlnProArgProSerLeuProLysSerThrMetH1 1670
4310 .....GCGTCCCGGCTCGTCTTTATGCT 4335
1670 sleuGlyProValValSerLysAlaLeuSerThrSerCysLeuValCys 1687
|||||
4336 CGAGGAGACCTGTGTGACAGAGCTTGGTTATGGGACACTGTTTGTCT 4385
1687 yLeuGlyGlnAsnProAlaAsnPheLysAspLeuGlyLysLeuGly 1703
|||||
4386 GTCTGTGTGACAGAGTGGCCATGACCGAGAGAGAGAGAGAGAGAGAG 4435
1704 ProTyrThrProGlnHis.....CysLeuProLysLysPro.. 1716
|||||
4436 CTTTTTATCCCAAGATTTATGACAGCCTCTCCGAGAGAGAGAGAGAG 4485
1717 .....LysLeuLysGlnLysValArgProGlnGly 1727
4486 TAGAGAGGCCACAGAAATGACAGAGCAAAATGATGAGGAGCAGC..... 4528
1727 hrCysGlnGlnAlaSerLeuProLeuGlnArgThrLeuLysGlyProG 1743
4529 .....AAAAGTCTTATGAGCTCCAGAGAGAGAGAGAGAGAGAGAG 4573
1743 lucysAlaAlaAlaAlaThrAlaGlyLysProPro 1754
4574 GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4608
seq_name: gb_sts:AF021118
seq_documentation_block:
LOCUS AF021118 495 bp DNA STS 03-APR-1998
DEFINITION Homo sapiens trinuclotide repeat ctg-37, sequence tagged site.
ACCESSION AF021118
VERSION AF021118.1 GI:3015591
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 495)
REFERENCE 1
Phillibert,R.A., Horelli-Kuitunen,N., Robb,A.S., Lee,Y.H.,
Long,R.T., Damschroder-Williams,P., Martin,B.M., Brennan,M.B.,
Palotie,A. and Gims,E.I.
The characterization and sequence analysis of thirty CTG-repeat
containing genomic cosmid clones
Eur. J. Hum. Genet. 6 (1), 89-94 (1998)
JOURNAL MEDLINE 98454304
REFERENCE 2 (bases 1 to 495)
Phillibert,R.A., Horelli-Kuitunen,N., Robb,A.S., Lee,Y.H.,
Long,R.T., Damschroder-Williams,P., Martin,B.M., Brennan,M.B.,
Palotie,A. and Gims,E.I.
Direct submission
JOURNAL Submitted (28-AUG-1997) Clinical Neuroscience Branch, National
Institutes of Mental Health, 49 Convent Drive Rm B1E16, Bethesda,

```



```

|||||
300 TGTCCAGGCTGTGTGGCAGGAGGCGCCNACTCTTTAAAGATTGCACACA 251
242 ProthralaglnProHlsAspargProleuthrAlaSerSerSerleuAl 238
250 CCGACTGCCCCAGCCCATGACAGCGCGCTGACTGCCAGCTCCAGCTGGC 201
258 AProglglnArGValGlnslnleuHlsAlaTyrGlnSerGlyArgLeuS 275
200 CCGGAGGAGCAGCGGCTCCAGATCTTCACTGCTACCTGCGGCGCTCA 151
275 eRTYrAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 291
150 GCTATGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGA 101
292 AlaIeuglnSerArgHlsAlaGlnGlnGlnGlnGlnGlnGlnGlnGln 308
100 GCCCTTACAGCGCGCCACATGCCAGGAAACCTCCATACCAAAACCT 51
308 uAlaIySTyGlnHlsTyGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 324
50 CGCCAACTATACAGCACTACGAGGCGAGGCGAGGCTACTGCAAGCG 2
seq_name: gb_hhg:AC026106

```

```

seq_documentation_block:
LOCUS AC026106 164485 bp DNK HTG 06-NOV-2000
DEFINITION Homo sapiens chromosome 12 clone RP11-527C10, WORKING DRAFT
SEQUENCE SEQUENCE, 16 unordered pieces.
ACCESSION AC026106 GI:11094470
VERSION AC026106.12 GI:11094470
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 164485)
REFERENCE
AUTHORS

```

1 (bases 1 to 164485)  
 Wuzy,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
 Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbatis,J.,  
 Benton,J., Bimage,K., Blankenburg,K., Bonini,D., Bouck,J.,  
 Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Bunney,C.,  
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Cartron,T.F.,  
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
 Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,  
 Coyle,M.D., Dethorne,S.R., David,R., Davila,M.L., Davis,C.,  
 Day-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
 Elnaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,  
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,  
 Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,  
 Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A.,  
 Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holikway,C.,  
 Hollins,B., Homs,F., Howard,S., Huber,J., Huliy,S., Hume,J.,  
 Jackson,L.F., Jacobson,R., Jia,Y., Johnson,R., Jollivet,S.,  
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korah,J.,  
 Kovar,C., Kravovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,  
 Lewis,L., Li,J., Li,Z., Lichtenage,O., Lieu,C., Liu,J., Liu,W.,  
 Loulsegad,H., Lozado,R.J., Lu,K., Lucier,A., Lucier,R., Luna,R.,  
 Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,  
 Martinez,E., Massey,E., Mawhney,E., Mcleod,M.P., Meador,M.,  
 Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabati,K.,  
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,S., Nwokwenwo,S.,  
 Nguyen,A., Nguyen,N., Nguyen,N., Oviado,R., Pace,A., Payton,B.,  
 Ogutu,M., Okunoyu,G., Otaguye,N., Oviedo,R., Pichens,R., Pilius,E., Pu,L.L.,  
 Peery,J., Perez,L., Peters,L., Pickens,R., Pitman,E., Pu,L.L.,  
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojoubokan,I., Rolfe,M.,  
 Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N.,  
 Sisson,I., Sodergren,E., Soule,T., Sparks,A., Stenley,H.,  
 Stone,H., Sutton,A., Sytek,A., Tabor,P., Tamerisa,K., Tameris,K.,  
 Tang,H., Tansey,K., Taylor,C., Taylor,I., Telford,B., Thomas,R.,  
 Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,  
 Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,

COMMENT  
 REFERENCE  
 JOURNAL  
 TITLE  
 JOURNAL

Watlington,S., Williams,G., Williamson,A., Wleczek,R., Woden,S.,  
 Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
 and Gibbs,R.  
 Direct Submission  
 2 (bases 1 to 164485)  
 Worley,K.C.  
 Direct Submission  
 Submitted (19-MAR-2000) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Nov 4, 2000 this sequence version replaced gi.19966554.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 Project Information  
 ----- Project Information  
 Center project name: HMI  
 Center clone name: RP11-527C10  
 ----- Summary Statistics  
 Sequencing vector: M13; L08821  
 Chemistry: Dye-terminator Big Dye; 91% of reads  
 Chemistry: Dye-terminator Big Dye; 91% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 16300 bases at least Q40  
 Consensus quality: 156814 bases at least Q20  
 Estimated insert size: 154949; sum-of-contigs estimation  
 Quality coverage: 0x in Q20 bases; agarose-fp estimation  
 Quality coverage: 4.5x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_drift\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_drift_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 16 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 1  
 36628: contig of 36628 bp in length  
 36728: gap of unknown length  
 36729: contig of 31842 bp in length  
 36870: gap of unknown length  
 68571: contig of 25208 bp in length  
 68572: gap of unknown length  
 93879: contig of 17701 bp in length  
 93879: gap of unknown length  
 111679: contig of 12207 bp in length  
 111680: gap of unknown length  
 111780: contig of 12398 bp in length  
 123987: gap of unknown length  
 124087: contig of 9983 bp in length  
 134075: gap of unknown length  
 134175: contig of 7910 bp in length  
 142085: gap of unknown length  
 142185: contig of 7295 bp in length  
 142186: gap of unknown length  
 142186: contig of 3258 bp in length  
 149481: gap of unknown length  
 149481: contig of 3258 bp in length  
 152938: gap of unknown length  
 152938: contig of 3101 bp in length  
 156039: gap of unknown length  
 156039: contig of 1611 bp in length  
 156040: gap of unknown length  
 156040: contig of 1455 bp in length  
 157751: gap of unknown length  
 157751: contig of 1455 bp in length  
 159305: gap of unknown length  
 159305: contig of 1037 bp in length  
 160442: gap of unknown length  
 160442: contig of 1308 bp in length  
 160443: gap of unknown length  
 160443: contig of 1308 bp in length  
 161850: gap of unknown length  
 161850: contig of 1358 bp in length  
 163309: gap of unknown length  
 163309: contig of 1077 bp in length.  
 164485: contig of 1077 bp in length.

[illegible]



[illegible]





OM of: US-09-508-821b-6 to: N\_Geneseq\_1101.\* out\_format: pfs  
Date: Dec 4, 2001 2:04 PM  
About: Results were produced by the Gencore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:  
-MODE=frame+pn.model -DEV=xlh  
-O=us0950821.spool/us0950821/runat\_04122001\_114530\_10136/app-query.fasta.1.1842  
-B=N\_Geneseq\_1101 -OPMT=fastap -SUFFIX=ring -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPTOL=0.000 -LOOPTOL=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-EGAPOP=6.000 -EGAPEXT=7.000 -TGAPOP=10.000 -TGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blonum62  
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR\_SCORE=ptc  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFM=pts  
-NORM=ext HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0950821\_000001\_1.369 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XMPYX -WAIT -THREADS=1

Search information block:  
Query: US-09-508-821b-6  
Query length: 1755  
Database: N\_Geneseq\_1101.\*  
Database sequences: 930621  
Database length: 42862619  
Search time (sec): 174.910000

seq_list:	Strd Orig	ZScore	EScore	len	Documentation
Sequence					
/SID2/gcgdata/geneseq/NA1999.DAT:AA23983			9265.00	5639.23	4.9e-306
/SID2/gcgdata/geneseq/NA2000.DAT:AA288975			8479.00	5159.58	2.6e-279
/SID2/gcgdata/geneseq/NA2000.DAT:AA288975			1790.00	1095.30	6.2e-53
/SID2/gcgdata/geneseq/NA2001.DAT:AA288975			811.00	503.72	5.7e-20
/SID2/gcgdata/geneseq/NA2001.DAT:AA288975			811.00	503.72	5.7e-20
/SID2/gcgdata/geneseq/NA1994.DAT:AA288975			508.00	309.93	3.5e-09
/SID2/gcgdata/geneseq/NA1997.DAT:AA288975			493.00	292.81	3.1e-08
/SID2/gcgdata/geneseq/NA2000.DAT:AA288975			493.00	292.81	3.1e-08
/SID2/gcgdata/geneseq/NA1998.DAT:AA288975			491.00	289.35	4.8e-08
/SID2/gcgdata/geneseq/NA1999.DAT:AA288975			491.00	289.35	4.8e-08
/SID2/gcgdata/geneseq/NA1996.DAT:AA288975			470.00	289.24	4.9e-08
/SID2/gcgdata/geneseq/NA2000.DAT:AA288975			470.00	289.24	4.9e-08
/SID2/gcgdata/geneseq/NA1997.DAT:AA288975			468.50	279.03	1.8e-07
/SID2/gcgdata/geneseq/NA1998.DAT:AA288975			468.50	279.03	1.8e-07
/SID2/gcgdata/geneseq/NA2000.DAT:AA288975			468.00	273.01	3.9e-07
/SID2/gcgdata/geneseq/NA2001.DAT:AA288975			464.50	269.87	8.6e-08
/SID2/gcgdata/geneseq/NA2000.DAT:AA288975			462.50	268.36	7.1e-07
/SID2/gcgdata/geneseq/NA1997.DAT:AA288975			450.50	264.20	1.2e-06
/SID2/gcgdata/geneseq/NA1997.DAT:AA288975			450.50	264.20	1.2e-06
/SID2/gcgdata/geneseq/NA2001.DAT:AA288975			447.50	261.87	1.6e-06
/SID2/gcgdata/geneseq/NA2001.DAT:AA288975			445.50	261.72	6.8e-07
/SID2/gcgdata/geneseq/NA2000.DAT:AA288975			445.50	261.72	6.8e-07
/SID2/gcgdata/geneseq/NA2000.DAT:AA288975			445.50	261.17	1.5e-06
/SID2/gcgdata/geneseq/NA2000.DAT:AA288975			445.50	261.17	1.5e-06
/SID2/gcgdata/geneseq/NA2000.DAT:AA288975			445.50	261.08	1.6e-06
/SID2/gcgdata/geneseq/NA2000.DAT:AA288975			445.50	261.08	1.6e-06
/SID2/gcgdata/geneseq/NA1997.DAT:AA288975			442.50	259.82	8.7e-07
/SID2/gcgdata/geneseq/NA1997.DAT:AA288975			442.50	259.82	8.7e-07
/SID2/gcgdata/geneseq/NA2000.DAT:AA288975			438.50	259.29	2.3e-06
/SID2/gcgdata/geneseq/NA2000.DAT:AA288975			438.50	259.29	2.3e-06
/SID2/gcgdata/geneseq/NA1998.DAT:AA288975			433.50	253.88	4.6e-06
/SID2/gcgdata/geneseq/NA1998.DAT:AA288975			433.50	253.88	4.6e-06
/SID2/gcgdata/geneseq/NA1997.DAT:AA288975			429.50	251.40	6.4e-06
/SID2/gcgdata/geneseq/NA1997.DAT:AA288975			429.50	251.40	6.4e-06
/SID2/gcgdata/geneseq/NA1997.DAT:AA288975			426.50	249.60	7.9e-06
/SID2/gcgdata/geneseq/NA1997.DAT:AA288975			426.50	249.60	7.9e-06
/SID2/gcgdata/geneseq/NA2001.DAT:AA288975			422.50	243.34	1.8e-05
/SID2/gcgdata/geneseq/NA2001.DAT:AA288975			422.50	243.34	1.8e-05
/SID2/gcgdata/geneseq/NA2001.DAT:AA288975			421.50	237.85	2.7e-06
/SID2/gcgdata/geneseq/NA2001.DAT:AA288975			421.50	237.85	2.7e-06
/SID2/gcgdata/geneseq/NA2001.DAT:AA288975			421.50	245.79	1.3e-05

seq\_documentation\_block:  
ID: AA23983 standard; DNA: 6022 BP.  
AC: AA23983;  
XX: 23-JUN-1999 (first entry)  
XX: Human hGRI DNA.  
XX: Polymorphic CAG repeat; hGRI; diagnosis; prognosis; schizophrenia;  
XX: human; transcription factor; neuroleptic activity; affective disorder;  
XX: manic depression; neurodevelopmental brain disease; detection;  
XX: phenotypic variability; ss.  
XX: Homoc sapiens;  
XX: W09915639-AL.  
XX: 01-APR-1999.  
XX: 18-SEP-1998; 98WC-CA00884.  
XX: 19-SEP-1997; 97CA-2216057.  
XX: (GPMC) UNIT MCGILL.  
XX: Benkelat C, Joobier R, Rouleau GA;  
XX: WPI; 1999-254703/21.  
XX: A human GRI gene containing a transcribed polymorphic CAG repeat,  
XX: useful in the diagnosis and treatment of schizophrenia  
XX: Claim 1; Fig 4A-E; 41pp; English.  
XX: This invention describes novel human GRI (hGRI) transcription factor  
XX: gene with neuroleptic activity containing a transcribed polymorphic  
XX: CAG repeat. Allelic variants of the hGRI gene CAG repeat are associated  
XX: with schizophrenia, affective disorders (especially manic depression),  
XX: neurodevelopmental brain diseases or with phenotypic variability with  
XX: respect to long term response to neuroleptic medication. Short  
XX: (171-177 bp) allelic variants of the hGRI gene, are  
XX: indicative of non-severe schizophrenia and neuroleptic response in  
XX: patients. Probes and/or primers designed using the hGRI gene can be  
XX: used to identify genes interacting with a biochemical pathway affected  
XX: by the hGRI gene. The identified gene role can then be evaluated in  
XX: psychiatric patients. Therapeutic agents can be identified by  
XX: administering the agent to a transgenic mammal (or schizophrenic  
XX: patients) and evaluating the prevention and/or treatment of development  
XX: of schizophrenia.  
XX: Sequence 6022 BP; 1247 A; 2060 C; 1799 G; 916 T; 0 other;  
XX: S2  
XX: alignment\_scores:  
XX: Quality: 9265.00 Length: 1755  
XX: Ratio: 5.281 Gaps: 0  
XX: Percent Similarity: 100.000 Percent Identity: 100.000  
XX: alignment\_block:  
XX: US-09-508-821b-6 x AA23983  
XX: Align seg 1/1 to: AA23983 from: 1 to: 6022  
XX: 1 MetGlnSerPheAaGlnArgCysGlyPheHisClyGlyGlnGlnAsntry 17

|||||  
490 ATGCACTCTTTTCGAGAAAGGTGTGTTTCCATGCGCAACACAGAACTA 539  
17 rGlnGlnThrSerGlnGlnThrSerArgLeuGluAsnTyrArgGlnPro 34  
|||||  
540 CCAGCAGACCTCGCAGAGAAACATCAGCGCTAGAGATTACAGCAGCGGA 589  
34 erGlnAlaGlyLeuSerCysAspArgGlnArgLeuLeuAlaLysAspTyr 50  
|||||  
590 GTGAGCGCGGGCTAAGCTGCGACCGCGGCGCGGCTGCTGCGCAAGACTAT 639  
51 TyrAsnProGlnProTyrProSerTyrGlnGlyValAlaGlyThrProSe 67  
|||||  
640 TATAACCGCCAGACCTTACCGAGCTATGAGGCTGGGCTGGCAGCGCTC 689  
67 rGlyThrAlaAlaAlaValAlaAlaAspLysTyrHisArgGlySerLysA 84  
|||||  
690 TGGCAGCTGCGAGCGCGGTGGCCCGCCACCAAGTACACCGAGGCGCAAG 739  
84 LaLeuProThrGlnGlnGlnGlyArgProAlaPheProGlyTyr 100  
|||||  
740 CCTTGCCTCCACAGCAGCGCTCGAGGAGGCGCGCTTCCCTGGCTAC 789  
101 GlyValGlnAspSerSerProTyrProGlyArgTyrAlaGlyGluGluSe 117  
|||||  
790 GCGCTCCAGCAGCAGCGCCCTACCCAGCGCGCTATGCTGTGTAGGAGAG 839  
117 rLeuGlnAlaTTPGlyValProGlnProProProGlnProGlnProL 134  
|||||  
840 CCTTCAGGCTTGGGGGGCCCGCAGCCAGCCACCCCGCCAGCGCGCCGAC 889  
134 euProAlaGlyValAlaLysTyrAspGluAsnLeuMetLysLysThrAla 150  
|||||  
890 TACCTGCGAGGGGTGGCCAAAGTATGATGAGAGCTTGATGATGATGATG 939  
151 ValProProSerArgGlnTyrAlaGlnGlnGlyAlaGlnValProPheAr 167  
|||||  
940 GTGCCCCCAGCAGCAGTATGAGAGAGGCGCGCCAGGTGCTTTTCG 989  
167 gThrHisSerLeuHisValGlnGlnProProProGlnGlnProLeuA 184  
|||||  
990 GACTCACTCCCTGCACGTCCAGCAGCAGCCCGCGCCCGCCAGCGCCCTGG 1039  
184 LaTyrProLysLeuGlnArgGlnLysLeuGlnAsnAspIleAlaSerPro 200  
|||||  
1040 CATACCCCAAGTCCAAAGGCGAAGCTCGAGACGACATGCTCCCTCCCT 1089  
201 LeuProPheProGlnGlyThrHisPheProGlnHisSerGlnSerPhePr 217  
|||||  
1090 CTGCCCTTCCCGCAGGGTACCCACTTCTCAGCATTCGCCAGTCTCTCC 1139  
217 oThrSerSerThrTyrSerSerSerValGlnGlnGlyGlnGlyAlaH 234  
|||||  
1140 CAGCTCTCCACTACTCTCTCTCTCCAGGGGTGGTGGCGAGGGGGCCC 1189  
234 sSerTyrLysSerCysThrAlaProThrAlaGlnProHisAspArgPro 250  
|||||  
1190 ACTCCTATAGAGTTGCACAGCAGCCGACTGCGCCAGCCCGCCATGACAGCG 1239  
251 LeuThrAlaSerSerSerLeuAlaProGlyGlnArgValGlnAsnLeuH 267  
|||||  
1240 CTGACTGCCAGCTCCAGCTGCGCCCGCGGCGGCGGCTCCAGAACTTTCA 1289  
267 sAlaTyrGlnSerGlyArgLeuSerTyrAspGlnGlnGlnGlnGlnGln 284  
|||||  
1290 TGGCTACCATGCGGGCGCGCTCAGCTATGNCACAGCAGCAGCAGCAGC 1339  
284 InGlnGlnGlnGlnGlnGlnAlaLeuGlnInSerArgHisHisAlaGln 300  
|||||  
1340 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1389  
301 GluThrLeuHisTyrGlnAsnLeuAlaLysTyrGlnHisTyrGlyGlnG 317  
|||||

1390 GAAACCTCTCCATTACAAAACCTCGCCCAAGTATCAGCACTACGGGCGAGCA 1439  
317 nGlyInGlyTyrCysGlnProAspAlaAlaValArgThrProGluGlnT 334  
|||||  
1440 AGGCCAGAGGCTACTGCGCGGAGCGCAGCGCTCGGACCCCGCAGAGCAGT 1489  
334 yTyrGlnThrPheSerProSerSerSerHisSerProAlaArgSerVal 350  
|||||  
1490 ACTACAGAGCTTCCAGCCCGCAGCTCCAGCAGCTCAGCCCGCGCTCGGTG 1539  
351 GlyArgSerProSerTyrSerThrProSerProLeuMetProAsnLe 367  
|||||  
1540 GCGCGCTCAGCTTCTTACAGTTCACACCTCGCCGCTGATGCCAACT 1589  
367 uGluAsnPheProTyrSerGlnGlnProLeuSerThrGlyAlaPheProA 384  
|||||  
1590 GAGAACTTTCCTACAGCGCAGCGCTCAGCAGCGCGCTTCCCGGCTTCCCG 1639  
384 lagIyIleThrAspHisSerHisPheMetProLeuLeuAsnProSerPro 400  
|||||  
1640 CAGGATCACTGACACAGCCACTTCATGCCCTGCTCAATCCCTCCCGCA 1689  
401 ThrAspAlaThrSerSerValAspThrGlnAlaGlyAsnCysLysProLe 417  
|||||  
1690 ACGGATGCCAGCAGCTCTGTGGACACCCAGGCTGGCACTGCAAGCCCT 1739  
417 uGlnLysAspLysLeuProGluAsnLeuLeuSerAspLeuSerLeuGln 434  
|||||  
1740 TCAGAGAGCAGAGCTCCCTGAGAACCTCTGTGCGATCTCAGCTGCGAGA 1789  
434 erLeuThrAlaLeuThrLeuGlnValGluAsnIleSerAsnThrValGln 450  
|||||  
1790 GCCTCAGCGCGCTGACCTTACAGGTGGAGAACATCTCCAAACCGTCCAG 1839  
451 GlnLeuLeuLeuSerLysAlaAlaValProGlnLysGlyValLysAs 467  
|||||  
1840 CAGCTCTCTCTCAGAGCTGTGTGCCGCGAGAAAGGTTGCAAGAA 1889  
467 nLeuValSerArgThrProGlnHisLysSerGlnHisCysSerProG 484  
|||||  
1890 CCTCGTGTCCAGGACCCAGCAGCAGCATAAAGCCAGCAGCTGCAGCCCG 1939  
484 LuGlySerGlyTyrSerAlaGluProAlaGlyThrProLeuSerGluPro 500  
|||||  
1940 AAGGAGCGGCTACTCAGCGAGCGCCCGCAGGACACCTCGTGTGAGAGCG 1989  
501 ProSerSerThrProGlnSerThrHisAlaGluProGlnGlnAlaAspTy 517  
|||||  
1990 CCGAGCAGCAGCCACAGTCCAGCATCGGAGCGCGCAGGAGCGCGACTA 2039  
517 rLeuSerGlySerGluAspProLeuArgSerPheLeuTyrCysAsnG 534  
|||||  
2040 CCGTGGCGGCTCCAGAGCCACTGGAGCGCGCTTCTTACTGCAACC 2089  
534 InAlaArgGlySerProAlaArgValAsnSerAsnSerLysAlaLysPro 550  
|||||  
2090 AGGCCGCTGGCAGCCCTGCGAGGCTCAACAGCACTCGAGGCCAAGCC 2139  
551 GluSerValSerThrCysSerValThrSerProAspAspMetSerThrLy 567  
|||||  
2140 GAGTCCGTGTCCACTCTTGTGTGACCTCTCTGAGCATGTCCACCAA 2189  
567 sSerAspAspSerPheGlnSerLeuHisGlySerLeuProLeuAspSerP 584  
|||||  
2190 ATCTGAGGAGCTCTTCCAGAGCTACAGCGAGCTTCCGCTCGACAGCT 2239  
584 heSerLysPheValAlaGlyGluArgAspCysProArgLeuLeuLeuSer 600  
|||||  
2240 TCTCCAGTGTGTGGCGGTTGAGCGGGACTGTCCGCGGTGCTGCTCAGC 2289  
601 AlaLeuAlaGlnGluAspLeuAlaSerGlnIleLeuGlyLeuGlnGluAl 617  
|||||  
2290 GCCTGGCAGGAGGAGCCTGGCCCTCCGAGGATCTGGGCTGCAGAGAGC 2339

617 attleGIuValaAspLysAlaTrpAlaGluAlaProSerLeuValL 634  
|||  
2340 CATCGTGAAGAGCCGACAAAGCTGGCTGAGACACCAAGCCCTGGTCA 2389  
|||  
634 yaAspSerLysProPheSerLeuLysHisSerAlaCysLeu 650  
|||  
2390 AGACAGCAGCAGACCAACCTCTGCTGAGAACCAACAGCCCTGCTG 2439  
|||  
651 AspSerValAlaLysSerAlaTrpProArgProGluLysProGluAlaLe 667  
|||  
2440 GACTCTGTGGCAAGAGTGTGGCCCGGCTGGGAGCCGAGAGCCCT 2489  
|||  
667 uproAspSerLeuLysAspLysGlyLysAlaLysAspPheSerP 684  
|||  
2490 GCCGACCTCTTGCACCTGAGCAAGGGGCAAGCCCAAGGACTTCAACC 2539  
|||  
684 roGlyLeuPheGluAspProSerValAlaPheAlaTrpProAspProLys 700  
|||  
2540 CAGGGCTGTTTGAAGACCTTCGAGGCTTCTGACGCTGACCCCAAA 2589  
|||  
701 LysThrThrGlyProLeuSerPheGlyThrLysProThrLeuGlyValPr 717  
|||  
2590 AAGACACTGTCCTCTCTCTCTTGGTACCAAGCCCAACCTTGGGGTTC 2639  
|||  
717 oAlaProAspProThrThrAlaAlaPheAspCysPheProAspThrThra 734  
|||  
2640 TGCTCCAGACCCCACTACAGACCTTTGACTTGTCCCGAGACACACG 2689  
|||  
734 laAlaSerSerAlaAspSerAlaAspProPheAlaTrpProGluLysLeu 750  
|||  
2690 CTGCCACCTCAGGAGCAGCCCAACCCCTTGGCTTGGCCAGAGGAAAC 2739  
|||  
751 leuGIAspAlaCysProArgTrpGlyLeuHisProGlyLysLeuThrLys 767  
|||  
2740 CTGGGGATCTCTTCCAGGTGGGATGCAACCTGGCGAGCTTACCA 2789  
|||  
767 sGIyLeuGluGluGlyLysAlaAspArgLysLeuSerLysGlyAsp 784  
|||  
2790 GGGCTGAGCAGGAGGTGGAGGCTCAGATGACATCAGCAAGGAGAC 2839  
|||  
784 hHisGluAlaSerAlaCysLeuGlyPheGluGluLysProGlyLys 800  
|||  
2840 CCCATGAGGCTTGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2889  
|||  
801 GluLysValAlaSerLeuProGlyAspPheLysGluGluLysValGly 817  
|||  
2890 GAGAGGTGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2939  
|||  
817 yValLysGluGluAlaGlyLysLeuLysGlyProGluValAlaLys 834  
|||  
2940 GGTGAAGAGAGCAGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2989  
|||  
834 laAspArgTrpLeuGluAspSerArgHisCysCysSerThrAlaAspPhe 850  
|||  
2990 CTGACCGGTGGCTGAGAGACCGCAGCTCTCTTCCACCCGCACTTC 3039  
|||  
851 GIyAspLeuProLeuLeuProProThrSerArgLysGluLysLeuGluAl 867  
|||  
3040 GGGGACCTCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3089  
|||  
867 agLugluLysSerSerLeuCysGluLeuLysGlySerProGluGlu 884  
|||  
3090 TGAAGAGAGTACTCTCTCTATGATGATGATGATGATGATGATGATG 3139  
|||  
884 rGProGlyLeuGluAspProLeuSerProLysAlaProLeuLysCysThr 900  
|||  
3140 GGCCTGCACTGAGAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3189  
|||  
901 LysGluGluValGluGluValLeuAspSerLysAlaGlyTrpLysSer 917  
|||  
3190 AAGAGAGAGGTGAGAGGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 3239  
|||

917 oCysHisLeuSerGlyLysValIleLeuLysGlyProThrValGlyTr 934  
|||  
3240 GTGCCACCTCTCAGGGAGTCTCTCTCTCTCTCTCTCTCTCTCTCT 3289  
|||  
934 hrgLysLysValGluSerTrpPheGluLysSerSerLeuHisLys 950  
|||  
3290 CCGAGTCAAGGTCCAGAGCTGTTGAGTCTCTCTCTCTCTCTCTCTCT 3339  
|||  
951 ProGlyGluGluGlyProAspGlyLysValGluAlaProLysPheSerTh 967  
|||  
3340 CCAGGTGAAGAGGGCTGATGGGAGCAGAGCTCCAGGGGATTCACAC 3389  
|||  
967 rSerAspAlaSerLeuAlaGluLysProAsnLysProAlaValProGlu 984  
|||  
3390 CTCCGAGCCTCTCTGCGCCAGAGCCCAACAGCTCTCTCTCTCTCT 3439  
|||  
984 laProIleAlaLysLysGluProValProArgLysSerLeuArgSer 1000  
|||  
3440 CGCCATCCAGAAAGAAAGCCCTGCAAGGGGCAAAAGCTTACAGAAC 3489  
|||  
1001 ArgArgValHisArgLysLeuProGluAlaGluAspSerProCysArgL 1017  
|||  
3490 CGTGGGTGACACCGGGGCTGCCGAGCCGAGGACTCCCATGCAAGG 3539  
|||  
1017 apProValLeuProLysAspLeuLeuProGluSerCysThrGlyPro 1034  
|||  
3540 ACCAGTGTGCCAAAGACCTTCTCTCTCTCTCTCTCTCTCTCTCT 3589  
|||  
1034 roGluGlyLysGluGluAlaGluAlaProGlyValArgLysAlaSerGlu 1050  
|||  
3590 CCAGAGGACATGAAAGGGCTGAGGCCCAAGCGGGGGCTCGGAA 3639  
|||  
1051 GIyLeuProArgMetCysThrArgSerLeuThrAlaLeuSerGluProAr 1067  
|||  
3640 GGGCTCCCAAGAGTGTACTCTCTCTCTCTCTCTCTCTCTCTCTCT 3689  
|||  
1067 gThrProGlyProProGlyLeuThrThrProAlaProAspLysL 1084  
|||  
3690 CAGCGCCGAGCCCGAGGCTGACACACCCCTGCAACCCCAAGCAAC 3739  
|||  
1084 euGIyLysGluArgAlaAlaPheLysSerGlyLysArgValGlyLys 1100  
|||  
3740 TGGGGGCAAGCAGAGCGGCTTCAATCGGCAAGGGGGAG 3789  
|||  
1101 proSerProLysAlaAlaSerSerProSerAsnProAlaAlaLeuProVa 1117  
|||  
3790 CCCACACCAAGCTGCTCTCAGCCCAACCCGCGCTGCTGCTGCTGCT 3839  
|||  
1117 laAspSerSerProMetGlySerLysThrLysGluThrAspSerP 1134  
|||  
3840 GGCTCCGACAGACCCGATGGGCTCAGAGCCAGAGAGAGAGACTCAC 3889  
|||  
1134 roSerThrProGlyLysAspLysArgSerMetIleLeuArgSerArgThr 1150  
|||  
3890 CCAGACGCTGGCAAGAGCAGCAGCTCATGATCTCTCGGTCACCAAC 3939  
|||  
1151 LysThrGluGluIlePheHisSerLysArgArgArgProSerGluLys 1167  
|||  
3940 AAAACCCAGAGATCTTCCACTCCAAACCGCGAGAGCCCTCTGAGG 3989  
|||  
1167 gluProAsnCysArgAlaThrLysLysLeuAspAsnSerHisLeuP 1184  
|||  
3990 GCTCCCAACGCGGTGCCACAAAGAGCTCTCTCTCTCTCTCTCTCT 4039  
|||  
1184 roAlaThrPheLysValSerSerProGluLysGluLysValSer 1200  
|||  
4040 CCGCACTTCAAGGTCTCAGACAGCCCAAGAGAGAGAGAGGTAGC 4089  
|||  
1201 GluArgAlaArgValProLysProGlyValGlyLysLysLeuSerAsp 1217  
|||  
4090 CAGCGGCAAGGTCCTCCAAACCTGTCAGGAGAGAGAGAGAGAGAG 4139  
|||  
1217 gProLeuHisAlaLeuLysArgLysSerAlaPheMetAlaProValPro 1234  
|||

```

|||||
1410 GCGCCCTCCATGCGCTCAAAAGGAAGTGGCGCTTCATGGCGCGCTGCCCA 4189
1234 hrLysLysArgAsnLeuValLeuArgHisGlySerSerSerSerSerSerSer 1250
4190 CCAGAAGCGGAACCTGGCTTTGGCGGACCGGACGAGCAGCAGCAAC 4239
1251 AlaSerAlaMetGlyGluMetGlyArgArgArgGlyLeuArgValProPr 1267
4240 GCGAGTGAATGGGGAGATGGGAAGGAGGAGGAGGCGCTGAGGGTTCCGCC 4289
1267 oProSerSerArgGlyCysLeuProArgLysProSerProProArgA 1284
4290 ACCCTCTTCAAGAGATGCTTCTCCCAAGAAAGCCACGCCCCCAAGGG 4339
1284 laMetAlaSerLeuProGlnSerSerHisProArgArgProProPheLeu 1300
4340 CAATGGCGAGCTGCCACAAAGCTCCACGCCCGGAGAGCCCATCTCTG 4389
1301 ProGlnAlaArgLeuSerAlaAlaPheGlnGlyAlaMetLysThrLysVa 1317
4390 CCTCAAGCTCGCTCTGGCAGCTTCCAGGGGGCCATGAAGACCAAGGT 4439
1317 lLeuProProArgLysGlyArgGlyLeuLysLeuGluAlaIleValGlnL 1334
4440 GCTGCCACCCCGGAAGGCGGGGCGCTGAGCTGGAGGCCATCGTGCAGA 4489
1334 ysileThrSerProSerLeuLysLysPheAlaCysLysAlaProGlyAla 1350
4490 AGATCAGCTCGCCAGCTCAGAAAGTTGCGATGTAAGGCCGAGGGGCC 4539
1351 SerProGlyAsnProLeuSerProSerLeuSerAspLysArgGlyLe 1367
4540 TCTCTCTGTAATCTCTCAGGCCATCCCTTCCGACAAAGCCGTGGGCT 4589
1367 uLysGlyAlaGlySerProValGlyValGluGluGlyLeuValAsnV 1384
4590 CAAGGGTGC7GGGGGCGCCGCGGGGGTGGGAAGGCGCTGGTAAAG 4639
1384 alGlyThrGlyGlnLysLeuProThrSerGlyAlaAspProLeuCysArg 1400
4640 TGGGACCGGGGAGAGCTCCCACTCTGGGGCTGATCCGTATGCAGA 4689
1401 AsnProThrAsnArgSerLeuLysGlyLysLeuMetAsnSerLysLysLe 1417
4690 AATCCACCAACAGATCCTTAAAGGCAAACTCATGAACGTAGAAGCT 4739
1417 userSerThrAspCysPheLysThrGluAlaPheThrSerProGluAla 1434
4740 GTCTTCTACTGACTGTTTCAAAACCGAGGCGCTTCACTATCCCGGAGGCC 4789
1434 euGlnProGlyThrAlaLeuAlaProLysLysArgSerArgLysGly 1450
4790 TGCAGCTGGGGGAGTGGCTGGCGCTTAAAGAGGAGGAGCGGAAAGGC 4839
1451 ArgAlaGlyAlaHisGlyLeuSerLysGlyProLeuLysArgProTy 1467
4840 CGGCGAGGGGCCATGACTCTCCAAAGGCCCGCTGGAGAGCGGCCCTA 4889
1467 rLeuGlyProAlaLeuLeuThrProArgAspArgAlaSerGlyThrG 1484
4890 TCTTGGCGCGCTCTGCTCTGACTCCCGAGAGAGCGGCCAGTGGCACAC 4939
1484 lngLysAlaSerGluAspAsnSerGlyGlyGlyLysLysLysProLysMet 1500
4940 AAGGGCGCAGTGGGACACTCTGGTGGAGGAGGCAAGAAAGCAAGATG 4989
1501 GluLysLeuGlyProAlaSerGlnProProGluGlyArgProCysGlnPr 1517
4990 GAGGAGCTGGGCGCTCTCCACAGCCCGGAGGAGGCGCTGGCCAGCC 5039
1517 oGlnThrArgAlaGlnLysGlnProGlyHisThrAsnTySerSerTyrS 1534

```

```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ88975
seq_documentation_block:
ID AAZ88975 standard; cDNA; 5395 BP.
XX
AC AAZ88975;
XX
DT 26-MAY-2000 (first entry)
XX
DE Human HSET1 cDNA.
XX HSET1; human; chromosome 17; short arm; neuroprotection; diagnosis;
KW Smith Magenis Syndrome; multiple congenital anomaly; ss.
XX
OS Homo sapiens.

```

XX	
PN	MO200008143-A2.
XX	
PD	17-FEB-2000.
XX	
PF	06-AUG-1999; 99MO-DE02479.
XX	
PR	07-AUG-1998; 98DE-1035910.
XX	
PA	(DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PA	(PLAC ) MAX-PLANCK-INST MOLEKULARE GENETIK.
XX	
PI	Seranski P, Poustka A, Lehrnach H, Radelof U;
DR	WPI; 2000-195572/17.
DR	P-PsDB; AAy51611.
XX	
PT	Novel DNA encoding human HSGT1 useful for diagnosis or therapy of
PT	neurodegenerative disease, especially Smith Magenis Syndrome -
XX	
PS	Claim 1; Page 27-34; 39pp; German.
XX	
CC	This invention describes a novel human protein and its encoding nucleic acid, designated HSGT1 which has neuroprotective activity. The HSGT1 protein, or antibodies raised against it are useful as reagents to diagnose and/or for therapy of neurodegenerative disease, in particular Smith Magenis Syndrome which is a multiple congenital anomaly. This sequence encodes the human HSGT1 protein described in the invention
XX	
XQ	Sequence 5395 BP; 1117 A; 1877 C; 1595 G; 806 T; 0 other;

```

alignment_scores:
    Quality: 8479.00    Length: 1756
    Ratio: 5.189        Gaps: 6
Percent Similarity: 92.947    Percent Identity: 92.776

```

alignment\_block:

Align seg 1/1 to: AAZ88975 from: 1 to: 5395

1	MetGierPharqGluAqGcylpPhelIsgLYlVsgInglInuSPTy	17	
2	ATGCAAGCTTTCGAGAAAGATGTGTGTTCCATGCGAACAAGACTA	50	
3	17	rgInglInThrSerInglInuThSerArgLeuGluAnsInrArgInProS	34
4	51	CCAGAGACCTCGCAGGAACAATCAGCCCTAGAGAAATACAGCAACCGA	100
5	34	etGlnAlaGlyLeuSerCysAspArgInrArgInLeuAlaIysAspPyr	50
6	101	GTCAAGCGCGGGCTAAAGCTGGCAACCGCAGCGCTGCTGCCAAGACTT	150
7	51	TyrAsnProGlnProTyrProSerTyrGlnGlyAlaGlnGlyThrProSe	67
8	151	TATAACCGGACGCTTACCGGACGTATAGAGTGGCGCTGGCAAGCCCTC	200
9	67	rcLYThAlaAlaAlaValAlaAlaAspLysTyrHisArgIlySerLysA	84
10	201	TGGACATCGCAGCGCGGTGGCCCGCAAGTACCAACGAGCCAGCAAGG	250
11	84	IaLeuProThrArgInGlnGlyLeuGlnGlyArgProAlaPheProGlyTyr	100
12	251	CCCTGCCACACAGCAAGGCTTCGAGGGAGGCCGCTTCCCTGGCTAC	300
13	101	GlyValGlnAspSerSerProTyrProGlyArgTyrAlaGlyIngluSe	117
14	301	GGCGTCCAGGACGACAGCCCTACCAAGCCGCTTAGCTGGTGGAGAGG	350
15	117	rtGndAlaIaTPrpGlyAlaProGlnProGlnProGlnProGlnProGln	134
16	351	CCCTTAGAGCTTGGGGGGCCCAAGCAACACCCGCCCAAGCGCAAGCCAC	400

[illegible]

434 erLeuThrAlaLeuThrLeuGlnValGluAsnIleSerAsnThrValGln 450  
:::|||||  
1160 ACCTCAGCGGCTGACCTCAGAGTGGAGAACATCTCCACACCGTCCAG 1209  
451 GlnLeuLeuLeuSerLysAlaAlaValProGlnLysLysGlyValLysAs 467  
1210 CAGCTGCTGCTCTCCAGAGCTGCTGCTGCCGACAGAAAGGTGTCAAGA 1259  
467 nLeuValSerArgThrProGlnGlnHisLysSerGlnHisCysSerProG 484  
1260 CTTCTGTCTCAGAGCCCGACAGACATATAAAGCCAGCACTGCAGCCCG 1309  
484 luGlySerGlyTyrSerAlaGluProAlaGlyThrProLeuSerGluPro 500  
1310 AAGGAGCGGCTACTCAGCGGAGCCCGCAGGACACCGCTGCAGAGCGG 1359  
501 ProSerSerThrProGlnSerThrHisAlaGluProGlnGluAlaAspTy 517  
1360 CCGAGCAGCAGCCACAGTCCAGCATGCGGAGCGCCGAGAGCGCGACTA 1409  
517 rLeuSerGlySerGluAspProLeuGluArgSerPheLeuTyrCysAsnG 534  
1410 CTTGAGCGGCTCCGAGGACCACTGGAGCGCAGCTTCTCTACTGCAACC 1459  
534 lnAlaArgGlySerProAlaArgValAsnSerAsnSerLysAlaLysPro 550  
1460 AGGCCGCTGGCAGCCCTGCCAGGCTCAACAGCACTCGAAGGCGCAAGCCC 1509  
551 GluSerValSerThrCysSerValThrSerProAspAspMetSerThrIly 567  
1510 GAGTCCGCTGTCACCTGTTCTGTGACCTCTCTGAGCAGCATGTCCACCAA 1559  
567 sSerAspAspSerPheGlnSerLeuHisGlySerLeuProLeuAspSerP 584  
1560 ATCTGAGCACTCTCTCAGAGACCTACACGCGAGTGTGCGCTCGACAGCT 1609  
584 heSerLysPheValAlaGlyGluArgAspCysProArgLeuLeuLeuSer 600  
1610 TCTCCAGATTCTGCGCGGTGAGCGGACGTGTCGCGGCTGCTGCTCAGC 1659  
601 AlaLeuAlaGlnGluAspLeuAlaSerGluIleLeuGlyLeuGlnGluAl 617  
1660 GCCCTGGCAGAGAGAGCTTGGCTCCGAGATCTTGGGCTGCAGGAAGC 1709  
617 alleGlyLysAlaAspLysAlaThrPallGluAlaProSerLeuVal 634  
1710 CATCGTGGAGAGCGCACAAAGCTTGGGCTGAGCACCCAGCCTGTCTCA 1759  
634 yAspSerSerLysProProPheSerLeuGluAsnHisSerAlaCysLeu 650  
1760 AGGACAGCAGCAGCCACCTTCTGCTGGGAGAACACACAGCGCTGCTG 1809  
651 AspSerValAlaLysSerAlaThrProArgProGlyGluProGluAlaLe 667  
1810 GATCTGTGGCCAGAGTGGTGGCGCGCGCTGGGCTGGGAGCGCGAGGCC 1859  
667 uProAspSerLeuGlnLeuAspLysGlyClyAsnAlaLysAspPheSerP 684  
1860 GCCCGACTCTCTGACGTGGACAAAGGCGGCAATGCCAAGGACTTCAGCC 1909  
684 roGlyLeuPheGluAspProSerValAlaPheAlaThrProAspProLys 700  
1910 CAGGCTGTGTGAGACCTTCCGTGGCTTCCGTAGCCTGACCCGCA 1959  
701 LysThrThrGlyProLeuSerPheGlyThrLysProThrLeuGlyValP 717  
1960 AAGACACTGTGCTCTCTCTCTTGTGTACCAAGCCACCCCTTGGGCTTC 2009  
717 oAlaProAspProThrThrAlaAlaPheAspCysPheProAspThrThr 734  
2010 TGTCCAGACCCCACTACAGCAGCTTTTGTACTGTTCGCGACACAAACG 2059  
734 laAlaSerSerAlaAspSerAlaAsnProPheAlaThrProGluGluAsn 750

2060 CTGCCAGCTCAGCGGACAGCGCAACCCCTTTCCTGGCCAGAGGAAAC 2109  
751 LeuGlyAspAlaCysProArgTrpGlyLeuHisProGlyGluLeuThrLy 767  
2110 CTGGGGGATGCTTCTCCAGGTGGGATTCACCTGGCGAGCTTACCAA 2159  
767 sGlyLeuGlnGlnGlyLysAlaSerAspGlyLysLeuLysGlyAspT 784  
2160 GGGCTGGAGCAGGCTGGGAAGCCTCAGATGGCATCAGCAAGGGGACA 2209  
784 hrHisGlnAlaSerAlaCysLeuGlyPheGlnGluAspProProGly 800  
2210 CCTATGAGGCTTCGGCTTCCTGGCTTCAGAGGAGGACCCCTGGG 2259  
801 GluLysValAlaSerLeuProGlyAspPheLysGlnGlnGluValGlyCl 817  
2260 GAGAAGGTGGCTCTTCCCGGGGACTTCAAGCAGGAGGAGGTGGTGG 2309  
817 yValLysGlnGluAlaGlyLysLeuGlnCysProGlnValAlaLysA 834  
2310 GGTGAAGGAGGAGGAGGTGGGCTTGCAGTGGCCCGAGGTGGCCAAAG 2359  
834 laAspArgTrpLeuGluAspSerArgHisCysCysSerThrAlaAspPhe 850  
2360 CTGACCGGTGCTGGGAGGACAGCGGCTGCTGTTCACCGCGGACTTC 2409  
851 GlyAspLeuProLeuLeuProThrSerArgLysGluAspLeuGluAl 867  
2410 GGGGACCTCCACTGCTGCCACCCAGCAGGAGGAGGAGGAGGAGGAGC 2459  
867 aGluGluLysSerSerLeuGlyLeuLeuLeuGlySerProGluGlnA 884  
2460 TGAGGAGGAGTACTCTCTCTATGTGAGCTCTTGGGAGGAGGAGGAGGAG 2509  
884 rProGlyMetGlnAspProLeuSerProLysAlaProLeuLecCysThr 900  
2510 GGCTGGGATGCAGGAGCGGCTGTCCACCAAGCGCCACTCATCTGCACC 2559  
901 LysGluGluValGluGlnValLeuAspSerLysAlaGlyTrpGlySerPr 917  
2560 AAGGAGGAGTGGAGAGGTGTGGACTTCAAGCGCGGCTGGGCTCTCC 2609  
917 oCysHisLeuSerGlyGluSerValIleLeuLeuGlyProThrValGlyT 934  
2610 GTGCCACCTCTCAGGGAGTCCCTCATCTGCTGGGCGCTACAGTGGCA 2659  
934 hrGluSerLysValGlnSerTrpPheGluSerSerLeuSerHisMetLys 950  
2660 CCGAGTCAAGGTCCAGAGCTGTTTGTAGTCTCTCTGTACACATGAAG 2709  
951 ProGlyGluGlyProAspGlyGluArgAlaProGlyAspSerThrTh 967  
2710 CCAGTCAAGAGGGCTGTATGGGAGCGAGCTCCAGGGATTCACACAC 2759  
967 rSerAspAlaSerLeuAlaGlnLysProAsnLysProAlaValProGluA 984  
2760 CTCGGAGCGCTCTCTGGCCCAAGAGCCCAAGAGCTGCTGTGCCGAGG 2809  
984 laProIleAlaLysLysGluProValProArgLysLysSerLeuArgSer 1000  
2810 CGCCCATCGCAAGAAGAGGCTGTGCCAGCGGGCAAAAGCTTACGGAGC 2859  
1001 ArgArgValHisArgGlyLeuProGluAlaGluAspSerProCysArgAl 1017  
2860 CGTCCGGTGCACCGGGGCTGCCGAGGCGAGGACTCCCTATGTCAGGGC 2909  
1017 aProValLeuProLysAspLeuLeuProGluSerCysThrGlyProP 1034  
2910 ACCAGTGTGCCCAAGAGCTTGTCTCCCTGAACTCTGTCACAGGGCCCC 2959  
1034 roGlnGlyGlnMetGluGlyAlaGlyAlaProGlyArgGlyAlaSerGlu 1050

2960 CCCAGGAGACAGATGAAAGGGGCTGAGAGCCAGCCGGAGGCGCTCGGAA 3009  
1051 G1yleuProArgMetCysThrArgSerLeuThrAlaLeuSerGluProAr 1067  
3010 GGAGCTCCAGAGATGTACTGCTCTCTCCAGCCCTGAGAGAGCCCG 3059  
1067 gthrProGlyProGlyLeuThrThrProAlaProPheAspLysL 1084  
3060 CAGCCCGGAGCCCGAGGCTGACACACCCCTGCAACCCCGAGCAAC 3109  
1084 euGlyGlyLeuGlnArgAlaAlaPheLysSerGlyLysArgValGlyLys 1100  
3110 TGGGGGGGCAAGACGAGCGGCTTCAAGTGGGCAAGGGGTGGGAG 3159  
1101 ProSerProLysAlaAlaSerSerProSerAsnProAlaLeuProVa 1117  
3160 CCTCACCCAAAGCTGCTCCAGCCCGAGCAACCCGCGCCCTCCGT 3209  
1117 AlaSerAspSerSerProMetGlySerLysThrLysGluThrAspSer 1134  
3210 GGCTCCGAGACAGACCCGATGGGCTCCAAAGCAAGAGAGAGAGCTAC 3259  
1134 roSerThrProGlyLysAspGlnArgSerMetIleLeuArgSerArgThr 1150  
3260 CCAGCAGCGCTGGCAAGAGACAGCGCTCCATGATCTTCGATCAGCAGC 3309  
1151 LysThrGlnGluIlePheHisSerLysArgArgArgProSerGluLys 1167  
3310 AAACCCAGAGATCTTCCATCCAGCGCGAGCGCCCTCTAGGGCGG 3359  
1167 gLeuProAsnGlyArgAlaThrLysLysLeuAspAsnSerHisLeuP 1184  
3360 GCTCCCACTGCGTGGCCAGCAAGAGCTCTCCACAAACCACTTCC 3409  
1184 roAlaThrPheLysValSerSerSerProGlnLysGluLysArgValSer 1200  
3410 CCGCCACATTCAGAGTCTCCAGAGCCCGAGAGAGAGGAGGAGTACG 3459  
1201 GlnArgAlaArgValProLysProGlyAlaGlySerLysLeuSerAspAr 1217  
3460 CAGCGGGGAAAGGTTCCCAACCTGGTGCAGGAGAGAGCTCTGACCG 3509  
1217 gProLeuHisAlaLeuLysArgLysSerAlaPheMetIaProValPro 1234  
3510 GCCCTCCATGCGCTCAAGAGAAATCGGCTTCATGGCGGCTCCCA 3559  
1234 hrlLysLysArgAsnLeuValLeuArgHisGlySerSerSerSerAsn 1250  
3560 CCAAGAGCGGAGACTGCTGCTTCGAGAGCGGAGAGAGAGAGAGAGC 3609  
1251 AlaSerAlaMetGlyLysMetGlyArgArgArgGlyLeuArgValProp 1267  
3610 GCCAGTGGCAATGGGGAGATGGGAAGAGAGAGAGAGCTGAGAGGTCCC 3659  
1267 roProSerSerArgGlyCysLeuLeuProArgLysProSerProArg 1283  
3660 CACCTCTCAAGAGAGATGCTTCCCAAGAAAGCCAGCCAGCCAGAG 3709  
1284 AlaMetAlaSerLeuProGlnSerSerHisProArgArgProPheL 1300  
3710 GCATGGGAGAGCTGCAAAAGCTCCACCCCGGAGAGACCCGATGCC 3759  
1300 euProGlnAlaArgLeuSerAlaAlaPheGlnGlyAlaMetLysThrLys 1316  
3760 TGGCTCAAGCTGCGCTCGGAGAGCTTCCAGGGGGCCATGAAACCA 3809  
1316 sValLeuProProArgLysGlyArgGlyLysLysLeuGlnAlaIleValG 1333  
3810 GGTGTGCAACCCCGAGAGGGCGGCGCTGAGCTGAGAGGCAATGTC 3859  
1333 LysLysLysSerProSerLeuLysLysPheAlaCysLysAlaProGly 1349  
3860 AGAAGATCACTCGCCAGCTCAAGAGATTCGATGAAAGGCCAGG 3909  
1350 AlaSerProGlyAsnProLeuSerProSerLeuSerAspLysArgL 1366  
3910 GCTCTCTCTGTATCTCTGAGCCCTTCTCCAGCAAGAGCCG 3959  
1366 LysLysGlyAlaGlyLysSerProValGlyValGluGlyLeuVala 1383  
3960 GCTAAGGTGCTGGGGGAGCCCACTGGGGGTGGAAAGAGCGTGTAA 4009  
1383 snValGlyThrGlyGlnLysLeuProThrSerGlyAlaAspProLys 1399  
4010 ATGGGGGACCGGGAGAAAGCTCCAACTTCGGGGCTGATCCGTATTC 4059  
1400 ArgAsnProThrAsnArgSerLeuLysGlyLysLeuMetLysSerLys 1416  
4060 AGAATCCAAACCAACAGATCTTAAAGGAGAACTCAAGAAC 4101  
1416 sLeuSerSerThrAspCysPheLysThrGlnAlaPheThrSerProGln 1433  
4101 4101  
1433 LeuGlnProGlyGlyThrAlaLeuAlaProLysLysArgSerArgLys 1449  
4101 4101  
1450 GlyArgAlaGlyAlaHisGlyLeuSerLysGlyProLeuGluLysArgPr 1466  
4101 4101  
1466 ctyLeuGlyProAlaLeuLeuThrProArgAspArgAlaSerGlyT 1483  
4102 4102  
1483 hrlGlnGlyAlaSerGluAspAsnSerGlyGlyGlyLysLysProLys 1499  
4115 CAAAGGGGCGAGAGAGCAACTGCTGAGAGAGAGAGAGAGAGAG 4164  
1500 MetGlnGluLeuGlyProAlaSerGlnProProGlnLysArgProCysG 1516  
4165 ATGAGAGAGCTGGGCTGGGCTCCAGCCCGGAGGGGAGGCGCTGCA 4214  
1516 ProGlnThrArgAlaGlnLysGlnProGlyHisThrAsnLysSer 1533  
4215 GCCCGAGACAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4264  
1533 YrsLysArgLysArgLeuThrArgGlyArgAlaLysAsnThrThrSer 1549  
4265 ATTCAAGCGGAGAGCGCTCTACCTGAGGCGGAGAGAGAGAGAGAG 4314  
1550 SerProCysLysGlyArgAlaLysArgArgArgGlnGlnGlnValLeuPr 1566  
4315 TCACCTCTAAGGGGCTGCCAAGCAGAGAGAGAGAGAGAGAGAGAG 4364  
1566 OleuAspProAlaGluProGlnLysLeuArgLysLysLysSerCysL 1583  
4365 CCGTGAATCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4414  
1583 YrsArgLeuArgSerAspSerArgThrProAlaPheSerProPheValArg 1599  
4415 AGCGGTGAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4464  
1600 ValGluLysArgAspAlaPheThrThrIleCysThrValValAsnSerPr 1616  
4465 GNGAG 4514  
1616 oGlyAspAlaProLysProHisArgLysProSerSerSerAlaSerSers 1633  
4515 TGAAGATGCGGCAAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAG 4564  
1633 eSerSerSerSerSerPheSerLeuAspAlaAlaCysLysAlaSerLeuAla 1649  
4565 CCGTCT 4614





1545 YSAsnThrThrSerSerProCysLysLysValArgAlaLysArgArgGln 1561  
 401 AGAAGACACACCTCTCACCTGTAAAGGGGGGCGCCACAGCGACGACGAG 450  
 1562 GlnGlnValLeuProLeuAspProAlaGlnProGlnIleArgLeuLysTrp 1578  
 451 CAGCAGGTGTGGCCCTGGATCCCGACAGACCTGAAATCCGCTCAAGTA 500  
 1578 TrpSerSerCysLysArgLeuArgSerAspSerArgTrpAlaPheS 1595  
 501 CATTCCTCTTGCAGAGGGCTGAGTGCAGACCGGACCGCCGCTTCT 550  
 1595 ePProPheValArgValGlnLysArgAspAlaPheThrTrpLeuCysThr 1611  
 551 CACCTTCGTGGGGGAGAGAGACGACGCGTTCACCACTATAGCACT 600  
 1612 ValValAsnSerProGlnAspAlaProLysProHisArgLysProSerS 1628  
 601 GTGTCACTCCCTCGAGATGGCCCAAGCCCAAGAGAACCTTCTCTC 650  
 1628 rSerAlaSerSerSerSerSerSerSerSerPheSerLeuAspAlaAla 1645  
 651 CTCTGCT 700  
 1645 LysAlaSerLeuAlaThrLeuProGlnLysSerLysLeuGlnProArgPro 1661  
 701 GGGGCTCCCTGGCCACACTCCGAGAGGCTCCATCCGACGCGGCGCC 750  
 1662 SerLeuProLeuSerSerThrMetHisLeuGlnProValAlaSerLysAl 1678  
 751 TCCTTGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 800  
 1678 AlaSerThrSerCysLeuValCysCysLeuGlnAsnProAlaAsn 1695  
 801 CCGAGTACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 850  
 1695 helysAspLeuGlnAspLeuCysGlyProTyrTyrProGlnHisCysLeu 1711  
 851 TCAAGACACTGGGGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
 1712 ProLysLysLysProLysLeuLysGlnLysValArgProGlnGlnLys 1728  
 901 CCAAAAGAGAGCAAACTCAAGAGAAAGTGGCGCCAGAGACGACCG 950  
 1728 sGlnGlnAlaSerLeuProLeuGlnArgTrpLeuLysGlnProGlnCys 1745  
 951 TATGAGAGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1000  
 1745 LalaAlaAlaThrAlaGlyLysProArg 1755  
 1001 CAGCTGCCGCCACTGCCGGGAGGCCGCCGAGG 1032  
 seqname: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAD05635  
 seq\_documentation\_block:  
 ID AAD05635 standard; CDNA: 2409 BP.  
 AC AAD05635;  
 DT 17-JUL-2001 (first entry)  
 DE Human secreted protein-encoding gene 16 CDNA clone HDPXY01, SEQ ID NO:67.  
 XX  
 XX Human: secreted protein; proliferative disorder; cancer; tumour;  
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
 KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;  
 KW endocrine disorder; infection; wound healing; vulnertary;

KW cell culture; chemotaxis; food additive;  
 KW binding partner identification; ss.  
 OS Homo sapiens.  
 TH Key location/Qualifiers  
 FT CDS 33..329  
 FT /tag= a  
 FT /product= "Human secreted protein"  
 FT sig\_peptide 33..143  
 FT /tag= b  
 FT mat\_peptide 144..326  
 FT /tag= c  
 FT /product= "Human mature secreted protein"  
 PN W0200134627-A1.  
 PD 17-MAY-2001.  
 PF 08-NOV-2000; 2000MO-US30628.  
 PR 12-NOV-1999; 99US-0164744.  
 PR 30-JUN-2000; 2000US-0215140.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Ruben SM, Komatsoulis GA, Baker KP, Young PE;  
 PI WPI: 2001-316491/33.  
 DR P-PSDB: AAE01826.  
 XX New nucleic acid molecules encoding human secreted proteins, used in  
 PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and  
 PT Parkinson's diseases and cancers -  
 XX  
 PS Claim 1; Page 465; 567pp; English.  
 CC AAD05579-AAD05658 represent cDNAs corresponding to 28 human secreted  
 CC protein genes and AAE01770-AAE01849 represent the proteins they encode.  
 CC AAE01850-AAE01860 represent human secreted protein fragments or variants.  
 CC The genes and their secreted proteins are useful for preventing,  
 CC treating or ameliorating medical conditions, e.g., by protein or gene  
 CC therapy. Pathological conditions can be diagnosed by determining the  
 CC amount of the new protein in a sample or by determining the presence of  
 CC mutations in the new genes. Specific uses are described for each of the  
 CC 28 genes, based on the tissues in which they are most highly expressed,  
 CC and include developing products for the diagnosis or treatment of  
 CC proliferative disorders, cancer, tumours, foetal and developmental  
 CC abnormalities, haematopoietic disorders, diseases of the immune system,  
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,  
 CC allergies, neurological disorders (e.g., Alzheimer's disease,  
 CC Parkinson's disease), cognitive disorders, foetal and developmental  
 CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,  
 CC cardiovascular disorders, angiogenic disorders, kidney disorders,  
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine  
 CC disorders, and infections. The proteins can also be used to aid wound  
 CC healing and epithelial cell proliferation, to prevent skin aging due to  
 CC sunburn, to maintain organs before transplantation, for supporting cell  
 CC culture of primary tissues, to regenerate tissues, to identify their  
 CC cognate ligands or binding partners, and in chemotaxis, and can be used  
 CC as a food additive or preservative to modify storage properties.  
 CC Antibodies specific for a protein of the invention can be used in  
 CC alleviating symptoms associated with the disorders mentioned above, and  
 CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked  
 CC immunosorbent assay (ELISA). The present sequence represents a human  
 CC secreted protein-encoding cDNA of the invention.  
 XX  
 XX Sequence 2409 BP; 490 A; 751 C; 711 G; 450 T; 7 other;

alignment\_scores:  
 Quality: 813.00 Length: 154  
 Ratio: 5.279 Gaps: 0

```

Percent similarity: 100.000 Percent Identity: 99.351
Lignment block:
US-09-508-821B-6 x AAD05635
Align seg 1/1 to: AAD05635 from: 1 to: 2409
1602 LysArgAspAlaPheThrThrLeuValValAsnSerProGlyAs 1618
10 CGCGGAGAGCGGTTCACCATATGCACTGTGTCAACTCCCTGGAGA 59
1618 pAlaProLysProHisArgLysProSerSerSerSerSerSers 1635
60 TGGGCCCAAGCCACACAGAGACCTTCCTCTGCTCTCTCTCTCAT 109
1635 erSerSerSerPheSerLeuAspAlaAlaLaglyAlaSerLeuAla 1651
110 CTTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 159
1652 ProGlyGlySerLeuLeuGlnProArgProSerLeuProLeuSer 1668
160 CTTGGAGGCTCATCTGCAGCGGGCCCTCTCTCTCTCTCTCT 209
1668 rMetHisLeuGlyProValValSerLysAlaLeuSerThrSerCys 1685
210 GATGCACCTTGGGGCTCTGTGTTCCAAAGCCCTGAGTACCTCT 259
1685 alCysCysLeuGlyGlnAsnProAlaAsnPheLysAspLeuGly 1701
260 TTTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 309
1702 CysGlyProLysTyrrProGluHisCysLeuProLysLysLys 1718
310 TGTGGGCTCTACTACCTGACACTGCTCCCAAGAGCAAACT 359
1718 uLysGluLysValArgProGluGlyThrCysGluGluAlaSer 1735
360 CAAGGAGAGGTCGGGCGCAGAGGACCTCTGTGAGAGGCTCT 409
1735 euGluAlaGlyThrLeuLysGlyProGluCysAlaAlaAlaThr 1751
410 TTGAGAGAACACTCAAGGTCGCGAGTGTGACGTCGCGCACT 459
1752 LysProProArg 1755
460 AAGCCGCCAGG 471
seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT: AAD05594
seq_documentation_block:
ID AAD05594 standard: cDNA; 766 BP.
AC AAD05594;
AC AAD05594;
DT 17-JUL-2001 (first entry)
DE Human secreted protein-encoding gene 16 cDNA clone HDPXY01, SEQ ID NO:26.
DE Human; secreted protein; proliferative disorder; cancer; tumour;
DE foetal abnormality; developmental abnormality; haematopoietic disorder;
DE immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
DE inflammation; allergy; neurological disorder; Alzheimer's disease;
DE Parkinson's disease; cognitive disorder; schizophrenia; asthma;
DE skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
DE cardiovascular disorder; angiotensin-related disorder; kidney disorder;
DE gastrointestinal disorder; pregnancy-related disorder; gene therapy;
DE endocrine disorder; infection; wound healing; vulnary;
DE cell culture; chemotaxis; food additive;
DE binding partner identification; ss.
OS Homo sapiens.
XX Key Location/Qualifiers

```

```

1603 ArgAspAlaPheThrTrpLeuValValSerProGlyAspAl 1619
1604 CGAGAGCGCTTCACCAACATATGCACTGTGTCACTCCCTGAAATG 52
1619 AProlyProHisArgLysProSerSerSerAlaSerSerSerSers 1636
1620 GCCCAAGCCGCCACAGAAAGCTTCCTCTGCTGCTCTCTCTCACTCT 102
1636 erSerSerPheSerLeuAspAlaAlaValSerLeuAlaThrLeuPro 1652
1637 CACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 152
1653 GlyGlySerLeuValSerLeuProArgProSerLeuProLeuSerThi 1669
1654 GGAGGCTCATCTCCAGACCGCGGCTCTCTCTCTCTCTCTCTCTCA 202
1669 ThrSerLeuProValSerLeuValSerLeuValSerLeuValC 1686
1687 GCACTGGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 252
1688 yScyLeuCyGluAsnProAlaAspPheLysAspLeuGlyAspLeuCy 1702
1689 GCTGCTCTCTGCAAAACCGCGCAACTCAAGACCTTGGGGAACCTCT 302
1703 GlyProTyrTyrProGlnHisCysLeuProLysLysLysProLysLeu 1719
1719 GGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 352
1719 sGlnLysValArgProGlnGlyThrCysGlnGlnLysLeuProLys 1736
1737 GGAGAGGCGGCGGCAAGAGGCACTGTGAGGAGGCGCTCTCTCTCT 402
1736 LurGThrLeuLysGlyProGlnCysAlaAlaAlaAlaAlaAlaGly 1752
1753 AGAGAACACTCAAGAGTCCGAGTGTGACCTCTCTCTCTCTCTCTCT 452
1753 ProProArg 1755
1754 CCCCCCAGG 461
seq_name: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.AA073500
seq_documentation_block:
ID AA073500 standard: DNA: 8438 BP.
AC AA073500:
XX
XX
XX 15-MAR-1995 (first entry)
XX
XX
XX DNA encoding Pseudorabies virus large latency transcript.
XX
XX Pseudorabies virus; PRV; LTV; large latency transcript;
KW attenuated virus; vaccine; early protein 0; BP0; HSV-1 ICP0;
XX protecting animals; deletion mutants; swine; ds.
XX
XX Pseudorabies virus.
XX
XX
XX Key Location/Qualifiers
XX misc_feature 1..7013
XX /tag= a
XX /note= "derived from PRV strain Inph"
XX misc_feature 7014..8425
XX /tag= b
XX /note= "derived from PRV strain Ka"
XX CDS 622..6498
XX /tag= c
XX /note= "encodes predicted amino acid sequence of ORF2"
XX TATA_signal 1..6
XX /tag= d
XX misc_feature 34
XX /tag= e
XX /note= "RNA cap site"
XX

```

```

FT polyA_signal 8382..8387
FT /tag= f
XX
XX US352596-A.
XX
XX 04-OCT-1994.
XX
XX 11-SEP-1992. 92US-0945283.
XX
XX 11-SEP-1992. 92US-0945283.
XX
XX (USDA ) US SEC OF AGRIC.
XX
XX Cheung AK, Wesley RD;
XX
XX MPI, 1994-316187/39.
XX
XX P-PSDB: AAR60620.
XX
XX New pseudorabies virus mutants for use in vaccine - having a
XX deletion and/or insertion in the early protein 0 gene or large
XX latency transcript gene
XX
XX Dislosure: Column 15-30; 43bp; English.
XX
XX AA073500 shows the pseudorabies virus (PRV) large latency transcript
XX (LTV). The basic sequence is derived from PRV strain Inph and PRV
XX strain Ka. The LTV overlaps and is transcribed in the opposite
XX orientation with respect to the BP0 (early polypeptide 0) and the
XX immediately early gene (IE180). BP0 is nonessential for replicatio,
XX LTV is the only gene expressed during PRV latency, and the IE180
XX gene is absolutely necessary for PRV replication. However there are
XX 2 copies of IE180 in the genome. It is expected that PRV lacking one
XX of the IE180 copies is viable. Deletions in the non-overlapping
XX regions of these 3 genes will generate single deletion mutants,
XX while deletions in overlapping regions will generate double deletion
XX mutants. The invention is concerned with the construction of attenuated
XX viruses which have a reduced ability to reactivate from latency. This
XX can be achieved by functionally disabling the expression of the BP0
XX gene, or by disrupting the synthesis of the LTV, or both. (See also
XX AA073501 and AAR60620-24)
XX
XX Sequence 8438 BP; 1141 A; 2916 C; 3327 G; 1054 T; 0 other;
XX
XX
XX alignment_scores:
XX Quality: 508.00 Length: 2060
XX Ratio: 0.647 Gaps: 104
XX Percent Similarity: 38.107 Percent Identity: 22.961
XX
XX alignment_block:
XX US-09-508-821b-6 x AA073500/rev ..
XX
XX Align seg 1/1 to reverse of: AA073500 from: 1 to: 8438
XX
XX 31 ArgGlnProSerGlnAlaGlyLeuSerCysAspArgGlnArgLeuAl 47
XX ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 7236 CGGCGCCATCTTGGCGCGGCGGCA.....AAGCGAGCGAGAGCA 7193
XX
XX 47 alyAspTyrTrpAsnProGlnProTyrProSerTyrGlnGly..... 61
XX ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 7192 AGCGGCGCAGAAAAATCCATTGGCGCGCGCGCGGAGAGTCCGCGGCA 7143
XX
XX 62 .....GlyAlaGlyThrProSerGlyThr 69
XX ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 7142 AATCGGCGCATTTGCTTACCTTACCTGGGCGGCGCTCTCTCTCTCTCT 7093
XX
XX 70 AlaAlaAlaValAlaAlaAspLysTyrHisArgGly..... 81
XX ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 7092 .....SerLysAlaLeuProThrGln..... 88
XX ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 7066 TTCACGTGGCGTGCAGGTACGAGCAGCATGTCTCTCTGCAACCGAGGGG 7017
XX

```

88 ..... 88  
7016 ATCCGATCGTCTCCGCTCCGCGCGGACTCTGAAGACTCCGGCTCCGG 6967  
89 .....GlnGlyLeuGlnArgProAlaPheProGly.....Ty 100  
6966 CGGTATCATCGCCTCGAGGAGCGCCGCCACACAGAGCTCTCGGCCAC 6917  
100 rGlyValGlnAspSerSerProTyrProGlyArgTyr..... 112  
6916 CAGAGAGAGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6867  
113 .....AlaGly.GluGlu.. 116  
6866 GCCGCCGCTCTCCGCGCTCGGCGCGCGCTCTCTCTCTCTCTCTCTCT 6817  
116 ..... 116  
6816 CCGATCGAGACCATGCGCGAGACTCTTTGACTTCATCTCAGACCGAGGG 6767  
117 .....SerLeuGlnAlaTyrpGlyAlaProGlnProProProGlnPro 131  
6766 CAACCTCAGCAGCTCTGG.....CGGCGCGCGCGCGCGCGCG 6726  
132 .....GlnProLeuProAlaGlyValAlaLysTyrAspGln 143  
6725 AGGAGAGAGGATCGCGCTCGGCGCGCGCGCGCGCGCGCGCGCG 6684  
143 uAsnLeuMetLysLysLysLysLysLysLysLysLysLysLysLys 159  
6683 .....GTCGCGCGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCT 6641  
159 LncGlnGlyAlaGlnValProPheArgThrHisSerLeuHisValGlnGln 175  
6640 CCGCGCGCGCGCTCTCTCG.....ACGAGC 6615  
176 ProProProProGlnGlnProLeuAlaTyrProLysLeuGlnArgGlnLy 192  
6614 CCG 6579  
192 sLeuGlnAsnAspIleAlaSerProLeuProPheProGlnGlyThrHisP 209  
6578 .....CGGAGCCACGAGCGCG..... 6561  
209 heProGlnHisSerGlnSerPheProThrSerSerThrTyrSerSerSer 225  
6560 ..... 6534  
226 ValGlnGlyGlyGlnGlyAlaHisSerTyr..... 236  
6533 GGCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6484  
237 LysSerCysThrAlaProThrAlaGlnProHisAspArgPro..... 250  
6483 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6434  
251 .....LeuThrAlaSerSerSerLeuAlaPro 259  
6433 TCTCCATCAGGCTCCGAGCAGCGTCTCTCTCTCTCTCTCTCTCTCT 6384  
260 GlyClnArgValGlnAsnLeuHisAlaTyrGlnSerGlyValGlnLeuSer 276  
6383 GGGCGG.....GTCGCGCGCGCGCGCG..... 6360  
276 rAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnAla 293  
6360 ..... 6360  
293 euGlnSerArgHisAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 309  
6359 ..CGACCGCGGAGCACTCGAGCGG..... 6336

340 LysTyrGlnHisTyrGlyGlnGln..GlyGlnGlyTyrCysGlnProAsp 325  
6335 .....CAGAGCGCGCGCGCGCGCGCGCGCGCGCTCTCTCTCTCT 6292  
326 AlaAlaValArgThrProGlnGlnTyrTyrGlnThrPheSerProSerSe 342  
6291 GCAGCC.....TCCCGCGCGCTC 6275  
342 rSerHisSerProAlaArgSerValGlyArgSerProSerTyrSerSer 359  
6274 CT 6225  
359 hrPro.....SerProLeuMet.....ProAsnLeuGluAsnPhe 370  
6224 CCGCGCGAGATGTCCTCCCGGAGGGGACCGCGTGTGAGCGCGCG 6175  
371 ProTyrSerGln.....GlnProLeuSerTh 379  
6174 CCGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6125  
379 rGlyAlaPheProAla..... 384  
6124 GCGCGAGCGCGCGCGCGCGCGCGCGCTCTCTCTCTCTCTCTCTCT 6075  
384 ..... 384  
6074 CCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6025  
385 .....GlyLeuThrAspHisSerHisPheMetProLeu..... 395  
6024 GACCGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5975  
396 .....LeuAsnProSerProThrAspAlaThrSerServ 407  
5974 CGCGCGCGGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5925  
407 alaThrGlnAlaGlyAsnCysLysProLeuGlnLysLysLysLeuPro 423  
5924 TCTCT..... 5920  
424 GluAsnLeuLeuSerAspLeuSerLeuGlnSerLeuThrAlaLeuThrLe 440  
5919 .....TCCACCCCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5882  
440 uGlnValGluAsnIleSerAsnThrValGlnGlnLeuLeuSerLysA 457  
5881 CGCGCTCGAGCC..... 5869  
457 laAlaValProGlnLysLysGlyValLysAsnLeuValSerArgThrPro 473  
5868 .....GTCGCGCGCGCGCGCG..... 5854  
474 GluGlnHisLysSerGlnHisCysSerProGlyGlySerGlyTyrSerAl 490  
5853 GAGAGAGAGAG.....GACGAGCTCGGCTCTCTCTCTCTCTCTCT 5822  
490 aGluProAlaGlyThrProLeuSerGluProProSerSerThrProGlnS 507  
5821 GGAC.....GGCGGGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 5782  
507 erThrHisAlaGluProGlnGlnAlaAspTyrLeuSerGlySerGluAsp 523  
5781 .....CGCGCGCGGAGAGGGC 5767  
524 ProLeuGlnArgSerPheLeuTyrCysAsnGlnAlaArgGly..... 537  
5766 GCGCT 5717  
538 .....SerProAlaArgValAsnSerAsnSerLysAla..... 548  
5716 CTCGGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5667  
549 .....LysProGlnSerValSerThrCysSerVal 558



1272 .....GlyCysLeuLeuProArg.....LysProSerProPhe 1282  
11046 tglValAserGluGlyLeuProArgMetCysThrArgSerLeuThrAla 1062  
3139 GGGCCCGGCTGCTTCGCCGCGTGAGGCTCTCTTCGCCCGCTGCAC 3090  
1282 oArgAlaMetAlaSerLeuPro.....Glns 1291  
3089 CGGGGCGGAGCGCTGCGCTCTCCGCCAGGACACGTGGCTACACG 3040  
1291 exSerHisProArgProPro.....Phe 1299  
3039 GTGAGCACCGCGCGCGCGCGGACCGCGCTCCGCTCCGCGCGCGC 2990  
1300 LeuProGlnAlaArgLeuSerAlaAlaPheGlnGlyAlaMetLysThrLy 1316  
2989 CTACCGGAGGCGCT..... 2975  
1316 sValLeuProArgLysGlyArgGlyLeuLysLeuGluAlaLeuValG 1333  
2974 .....GCTCCCGACCGTGGACGG.....CTGC 2953  
1333 InLysThrSerProSerLeuLysLysPheAlaCysLysAlaProGly 1349  
2952 AAGGACATGGCGCGCGCGCTCCGCGCTCGGGCT..CGGGGACCGCGA 2906  
1350 AlaSerProGlyLysProLeuSerProSerLeuSerAspLysAspArg 1366  
2905 CTTTGGACGGGG.....CGCG 2889  
1366 yLeuLysGlyAlaGlySerProValGlyValGluGluGlyLeuVala 1383  
2888 CTTTGGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2858  
1383 snValGlyThrGlyGlnLysLeuProThrSerGlyAlaAspProLeuLys 1399  
2857 .....GCT 2855  
1400 ArgAsnProThrAsnArgSerLeuLysGlyLysLeuMetAsnSerLysLy 1416  
2854 CGGCGCGCGCTCGCGCGCGCGCTTCGT..... 2828  
1416 sLeuSerSerThrAspCysPheLysThrGluAlaPheThrSerProGlu 1433  
2827 .....C 2827  
1433 laLeuGlnProGlyGlyThrAlaLeuAlaProLysLysArgSerArgLys 1449  
2826 TCTTCGGCGCGCGCGCGCTCGCGAGCTCCGCGCGCGCGCGCGCGCG 2777  
1450 GlyArgAla.....GlyAlaHisGlyLeuSerLys... 1459  
2776 GCGCGAGCTCGCGCGCTTCTCGCGCGCGCTGCTGCGCGCGCGCGCG 2727  
1460 .....GlyProLeuGluLysArgProTyrLeuGlyP 1470  
2726 AGCGCGCGCGCTGCTGCGCGCGCG.....CGCG 2696  
1470 roAlaLeuLeuThrProArgAspArgAlaSerGlyThrGlnGlyAla 1486  
2695 CTTCCG 2646  
1487 SerGluAspAsnSerGlyGlyGlyLysLysProLysMetGluGlu 1503  
2645 TCAGAG..... 2640  
1503 uGlyProAlaSerGlnProProGluGlyArgProCysGlnProGln...T 1519  
2639 .....CGAGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2594  
1519 hrArgAlaGlnLysGlnProGlyHisThrAsnTyrSerSerTyrSerLys 1535  
2593 CTTCCG 2562  
1535 hrArgAlaGlnLysGlnProGlyHisThrAsnTyrSerSerTyrSerLys 1548

```

2561 ..AGACTCTGCCGCGGTGAGATCTCGGCTCCGCGCGGCGAGCGC 2515
1549 .....SerSerProCysLysGlyArgAlaLysArgAla 1560
2514 CGGCCCCGCGCGCGCTCGCGGCCATCAAGCGAGGAGCGCATCTCGAGCA 2465
1560 rGlnGlnGlnValLeuProLeuAspProAlaGlnProGlnLysLeu 1576
2464 CGAGAGAG.....CCGAGGAGCGCGCGCAACCCCT..... 2436
1577 LysTyrLeuSerCysLysArgLeuArg..... 1586
2435 .....ACCTGCTGCTCCGCTGAGCGAGCGCGCGCGCGCGCT 2395
1587 .....SerAspSerArgTyrThrProAlaPheSerProPheValArgV 1600
2394 GGACTCTGACTCTGACTCTCGGCGCGCGCGCGCGCGCTCGA. 2346
1600 aGlnLysArgAspAlaPheThrThrLeuValValAsnSerPro 1616
2346 ..... 2346
1617 GlyAspAlaProLysProHisArgLysProSerSerSerAla...Serse 1632
2345 GGGCGCTCTCTCT...CCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 2239
1632 rSerSerSerSerSerSerPheSerLeuAlaAlaLysSerLeu 1649
2298 GTCCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2249
1649 lATHrLeuProGlyGly.....SerLeuGlnPro 1659
2248 CCTCGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2159
1660 ArgProSerLeuPro 1664
2198 CGGCGCGGTGCGCT 2184
seq_name: /SIDS2/gcgdata/geneseq/NA1997.DAT:AA789956
seq_documentation_block:
ID AA789956 Standard; DNA; 28958 BP.
XX
AC AA789956;
XX
DT 12-MAR-1998 (first entry)
XX
DE Sorangium cellulosum soraphen gene cluster genomic DNA.
XX
KW Soraphen; polyketide synthase; PKS; antipathogenic; antibiotic;
transgenic plant; phytopathogen; resistance; ss.
XX
OS Sorangium cellulosum.
XX
PN US562898-A.
XX
PD 02-SEP-1997.
XX
PF 01-JUN-1995; 95US-0457342.
XX
PR 20-AUG-1990; 90US-0570184.
PR 02-JUL-1992; 92US-0908284.
PR 31-AUG-1992; 92US-0937648.
PR 01-JUL-1993; 93US-0087636.
PR 08-JUN-1994; 94US-0258261.
XX
PA (CIBA ) CIBA GEIGY CORP.
XX
PI Beck JT, Gaffney RD, Hammer PE, Hill DS, Lam SF;
Ligon JM, Ryals JA, Schnupp T, Uknes SJ;
XX
DR WPI: 1997-447901/41.

```

```

XX
PR Protecting plants against pathogens with genetically transformed
PI biological control agent - which expresses all polypeptide(s)
XX involved in pyrrolnitrin biosynthetic pathway
XX
PS Ex 16; Column 104-128; 88pp; English.

```

This genomic DNA sequence encodes the Soraphen gene cluster. This sequence encodes two open reading frames (ORF's), the positions of which are not given in the specification. ORF1 is approximately 25.5 kb in size and encodes 5 biosynthetic modules homologous to the erythromycin genes of *Saccharopolyspora erythraea*. Each module contains a beta-ketoacyl synthase (KS), an acyltransferase (AT), a ketoreductase (KR) and an acyl carrier protein (ACP) domain. ORF2 is immediately adjacent to ORF1 and is thought to contain 3 modules. Soraphen is a type I polyketide synthase (PKS) which has antibiotic activity. Transgenic plants containing such antipathogenic genes like those encoded in the Soraphen cluster should have enhanced resistance to attack by phytopathogens.

XX  
SQ Sequence 28958 BP; 4084 A; 11863 C; 8356 G; 4655 T; 0 other;

alignment\_scores:  
Quality: 493.00 Length: 2095  
Ratio: 0.593 Gaps: 102  
Percent Similarity: 39.666 Percent Identity: 22.291

alignment\_block:  
US-09-508-821B-6 x AA789956

Align seg 1/1 to: AA789956 from: 1 to: 28958

```

62 GlyAlaGlyThrProSerGlyThrAlaAlaAlaAla.....Al 75
7509 GCGCGCGAGCGAGCGCGCTACGCGCGCGCGCGCGCTCTCTCTGACGC 7658
75 asPlyeTyrHisArgLysSerLysAlaLeuProThrGln..... 89
7559 CCGTCCGACACATCGCGCGAGCGCGCTGAGTACAGCGAGCGGTGGCGCT 7708
90 ..GlyLeuGlnGlyArgProAlaPheProGlyTyrGlyAlaGlnAspse 105
7709 GGGCGCGCTGCGCGCGCGCGCATGCGCGCATCAAGCGAGCGACGCCAC 7758
105 rSer.....ProTyrProGlyArgTyrAlaGlyGln 116
7759 CTCCCAAGCGCGCTCTGCGCGAGCGCGCGCTCTCTCTCTCTCTCTCTCT 7808
116 luserLeuGlnAlaTyrGlyAlaProGlnProProProGlnPro... 131
7809 GCTGCGCTGCTCTGAGACAGACAGACACGCTGACCGTGGCGACGA 7858
132 .....GlnProLeu...ProAlaGln 137
7859 TCGACTGGCGCGCTTTCGCGCTCTCTGAGCGCGCTCTCTCTCTCTCTCT 7906
137 yValAlaLysTyrAspGlnAsnLeuMetLysLysThrAlaValProProS 154
7907 .....CGCTCTCGCGCGCATTTGCCGCA 7928
154 erArgGlnTyrAlaGlnGlnValAlaGlnValProPheArgThrHisSer 170
7929 GCGCGAGCGCGCTCTGAGACAGCGAGCGCG..... 7960
171 LeuHisValGlnGlnProProProGlnGlnProLeuAlaTyrProly 187
7961 .....CGTCTCTGAGAGATGCGCGCGCG..... 7984
187 sleuGlnArgGlnLysLeuGlnAsnAspIleAlaSerProLeuProPheP 204
7984 ..... 7984

```



204 roGInglyThrHisPheProGlnHisSerGlnSerPheProThrSerSer 220  
 7985 .....  
 221 ThrTySerSerSerValGlnGlyGlyGlnGlyAlaHisSerTyry 237  
 7997 ACAAGCTCGGAGCGCTCGGAGAGGAGGAGCTTCGCTCGCTCGTCTCG 8046  
 237 sSerCys .....  
 8047 CTGGTGGCCACAGAGCGCTCGCTCGGCGCCAGGAGCGCTCGCA 8096  
 247 isAspArgProLeuThrAlaSerSerSerLeuAlaProGlyGlnArgVal 263  
 8097 TCTCGACCGGACAGAGGCTTCCTCGATC ..... 8125  
 264 GlnAsnLeuHisAlaTyrglnSerGlyArgLeuSerTyrglnGlnGln 280  
 8125 ..... 8125  
 280 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 297  
 8125 ..... 8125  
 297 isHisAlaGlnGlnThrLeuHisTyrglnAsnLeuAlaLysTyrglnHis 313  
 8126 ..... 8126  
 314 TyrGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 328  
 8144 TGGCGCTCGAGCTCGCGCGGCTTCACACAGGCGCCGCGATCAAGCTC 8193  
 328 lArgThrProGlnGln ..... 340  
 8194 CCGCGCACCTCGCTCGACCATCTCTCTCATCGAGTCGCGCTCT 8243  
 340 roSerSerHisSerProAlaArgSerValGlyArgSerProSerTy 356  
 8244 CTTGCGGAGCTCGCTCGCGGCGGCTTCGACAGGAGCTCTCGCTCG... 8290  
 357 SerSerThrProSerProLeuMetProAsnLeuGlnAsnPhe...ProTy 372  
 8291 ..... 8291  
 372 rSerGlnGlnProLeuSerThrGlyAlaPheProAlaGlyLeuThrAsp 389  
 8329 ACAGAGAGCGGCTCGCTCGGAGTGGCGCTCGGCTCGGCGGCG 8378  
 389 isSerHisPheMetProLeuLeuAsnProSer...ProThrAspAlaThr 404  
 8379 CGTGGCGAGTGGAGGCTCTTTGGAGTTCCTGGCCAGGAGCGGAGC 8428  
 405 SerSerValAspThrGlnAlaGlyAsnCysLysProLeuGlnLys...As 420  
 8429 GGCTCG ..... 8429  
 420 pLysLeuProGlnAsnLeuLeuSerAspLeuSerLeuGlnSerLeuThr 437  
 8455 TGGGATGCGCTCGCTACGACCGCGGAGCGGAGCGGAGCGGAG 8504  
 437 laLeuThrLeuGlnValGlnAsnIleSerAsnThrValGlnGlnLeu 453  
 8505 CTAGCTCGGCGGCTCGGCTCGGAGGAGGAGGAGGAGGAGGAGGAG 8554  
 454 LeuSerLysAlaAlaValProGlnLysLysGlyValLysAsnLeuVal 470  
 8555 CCTCTTTGGCATCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8589  
 470 rArgThrProGlnGlnHisLysSerGlnHisCysSerProGlnGlySer 487  
 8590 CTCAGCCCGC ..... 8590  
 487 lyTyrSerAlaGlnProAlaGlyThrProLeuSerGlnProPro... 501

8619 ARCTGCTGGCAGGCGCTCGAAGAGCGGCGGCTCGTCCCGCCACCTCA 8668  
 502 .....  
 8669 AGATTCCCGCCAGCGGCTTCGTCGCGCATCGCGCCAGCGAATACGCA 8718  
 509 lAlaGlu .....  
 8719 TTGCGAGAGGCGGAGCACCGAAGATTCCGACGCTATGCTCCCAAGGCA. 8767  
 516 spTyLeuSerGlySerGlnAspProLeuGlnArgSerPheLeuTyCys 532  
 8768 .....  
 533 AsnGlnAlaArgGlySerProAlaArgValAsnSerAsnSerLysAla 549  
 8804 CGTCTGGCTCGAAGGCGCG...CGCTCTGGTCGACACCGCTCGCTCC 8850  
 549 sProGlnSerValSerThrCysSerValThrSerProAspAspMetSer 566  
 8851 TCCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 8900  
 566 hrLysSerAspSerPheGlnSerLeuHisGlySerLeuProLeuAsp 582  
 8901 GTGCAACCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 8950  
 583 SerPheSerLysPheValAla ..... 589  
 8951 GCTTCGCT 9000  
 590 .....  
 9001 AAGACCTTCTCGGCGCACGCGCGCTAGGACGCGGAGAGGAGCGTCAT 9050  
 595 roArgLeuLeuLeuSerAlaLeu ..... 608  
 9051 CCGCT 9100  
 609 rGlnLeuLeuGlyLeuGlnGlnAlaGlyGlyGlyAlaAspLysAl 625  
 9101 TCCTCGCT 9150  
 625 a .....  
 9151 GGTATCACGCGCGCGCACGCGCTCCGACGAGAGTCTCTCGGCGCGC 9200  
 631 rLeuValLysAspSerSerLysProProPheSerLeu ..... 643  
 9201 GCTCCAGACGCGCGCATCACCGCGCGCGCGCTCGCTCGCTCGCTCG 9250  
 644 .....  
 9251 ATGCGACGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 9289  
 653 lAlaLysSerAlaThrProArgProGlyGlnProGlnAlaLeuProAsp 670  
 9290 .....  
 670 rLeuGlnLeuAspLysGlyAlaAlaLysAspPheSerProGlyLeu 686  
 9326 .....  
 687 PheGlnAspProSerValAlaPheAlaThrProAspProLysLysThr 703  
 9350 CGCTCAGACCAACATCGGCGCTCGGAGCGGCGCTCTCGGCTCGGCG 9399  
 703 rGlyProLeuSerPheGlyThrLysProThrLeuGlyValProAlaPro 720  
 9400 GTGCGCAAGATCG .....  
 720 spProThrThrAlaAlaPheAspCysPhePro .....  
 9443

9444 CACCTCCACAGGCGCCGCAAT...CCCTTGATGATGGATACA 9489  
733 ThrAlaIaSerSerAlaSerAlaSerProPheAlaTrpProGlu 749  
9490 CTGCGCATCGAGCTGTTGATACC...CCGAGGCTTGCGGCGCCCA 9533  
749 uAsnLeuGlyAspAlaCysProArgTrpGlyLeuHisProGlyLeuLeuT 766  
9534 CGAA...GATAGCATCCCGC... 9532  
766 hIlyseGlyLeuGluGluGlyLysAlaSerAspGlyIleSerIysGly 782  
9553 .....CGCGCGCGCTCCGCTTCGCACTCCGCGCAC 9588  
783 AspThrHis.....GluAlaSerAlaCysLeuGlyPheGlnI 795  
9589 AACGCCACGCTATCCTCGAGAGAGCTCCCGCGCCCTG...TC 9629  
795 uGluAspProProGlyGlyLysValAlaSerLeuProGlyAspPheLysG 812  
9630 GAGCGAGCGCGCACCTCAGACAGCGGCTCGGACCGCTCCCGCGCGCT 9679  
812 hngIuGluValGlyGlyValLysGluGluAlaGlyLysLeuLeuGlnCys 828  
9680 GTCCCGTGTCTGTCTGTGGCGAGAGCGCGCGCGCTCCCGCGCGAGCG 9729  
829 ProGlu.....ValAlaLysAlaAspArgTrpLeuGln 840  
9730 AAGCGGCTCGCGAGCACCTCTCCGCGACAGACCTCCGCGCTATACGA 9779  
840 pSerArgHisCysCysSerThr.....AlaAspPhe.GlyAspLeuPro 854  
9780 TGTGGCTATTCGAGGCGACACACCGCGCGCACTTCGAGTACCGCGCG 9829  
855 LeuLeuProProThrSerArgLysGluAspLeuGluAlaGluGluGly 871  
9830 CTCTCTCTGCGCGCGAGCGCGAGC..... 9853  
871 rSerSerLeuCysGluLeuLeuGlySerProGluGlnArgProGlyMetG 888  
9854 .AGCTCTCTCCGCGCTCGACTCGCTCGCGACAGACAGACCGC... 9895  
888 hAspProLeuSerProLysAlaProLeuIleCysThrLysGluGluVal 904  
9896 .....CCCCGAGCAGCGCTCTCGCGCGAGGAGGAGACCGC 9931  
905 GluGluValLeuAspSer.....LysAla.....GlyTr 914  
9932 GCAAGGTCTCTGCTCTTCCTTCGCGCAAGGCTCGCGAGGAGAGATG 9981  
914 pGlySerProCysHisLeuSerGlyLeuSerValIleLeuLeuGlyProT 931  
9982 GCCCTCTCCCTGCTGACTCTCGCGGCTTCGCGCTCAGTCAAGC 10031  
931 hValGlyThrGluSerLysValGlnSerTrpPheGluSerLeuSer 947  
10032 ATGCGAGCGCGCGCTCGCTCTCACTCG..... 10060  
948 HisMetLysProGlyGluGluGlyProAsp.....GlyGluArgAlaTr 962  
10061 .....AGTGAAGCTGCTCGCGCTCTCGCGCGAGGAGGCGC 10101  
962 oGlyAspSerThrThrSerAspAla..... 970  
10102 CCTCCCTCGAGCGCGCTCGAGCTGTCAGCGCGCGCTCTTGCGCTCAT 10151  
971 .....SerLeuAlaGluLysProAsnLysProAlaVal 981  
10152 GGTCTCCCTGCGCGCGCTCTGCGCTGCTGCGCTGAGCGCGCGCG 10201  
982 ProGluAlaProIleAlaLysGluProValProArgLysLysSer 998  
10202 TGTGTGGCGACAGCGCGAGATCCCGCGCGCTCTGTCGACAGGCGCT 10251

998 uArgSerArgArgValHisArgGlyLeuProGluAlaGluAspSerProC 1015  
10252 CTCTCCCTCGAGAGCGCGCGCGCATGCGCG...CCCT 10286  
1015 ys.....ArgAlaProValLeuProLysAspLeuLeuProGlu 1028  
10287 GCGAGGAAAGCGCTCACCGCGCGCGAGCGCGAGCGCGCGCGCG 10336  
1029 SerCysThrGlyProProGluGlyLys.....MetGluGlyAla.. 1041  
10337 TCGAGCTGCGCGCTCGAGCTCGAGACCTACCTGCTCCCTGCGCGAGC 10386  
1042 .....GlyAlaProGlyArgGlyAlaSerGln 1051  
10387 AGGCTCTCCACCGCGCGCTCAACAGCGCGCGCATCCCTGATTCGG 10436  
1051 LysLeuProArgMetCysThrArgSerLeuThrAlaLeuSerGluProArg 1067  
10437 CGAGCGCGCGCGCTCGAGCGCGCTGCTCGAGCTCTCAGCGCACAGG 10486  
1068 ...ThrProGlyProProGlyLeuThrThrThrProAlaPro..... 1080  
10487 TGTTCGCCCGCAAGATCCGCGTCACTACGCTCCACCTCCCGCAGATG 10536  
1081 .....ProAspLysLeuGly 1086  
10537 GAGCGCTCCAGAGAGACTCGCGCGAGCTCAACCTCTGCTCTCG 10586  
1086 LysGluAlaArgAlaAlaPheLysSerGlyLysArgValGlyLysProSer 1102  
10587 GAGCTCGAGCTCTCTCTTATT.....CGACCGTCA 10618  
1103 ProLysAlaIaSer...SerProSerAsnProAla..... 1113  
10619 CCGGACACAGGCTCGAGCGCGCTCGAGCTCGAGCGCGCTACGTATGCA 10668  
1114 .....AlaLeuProValAlaSerAspSerSerProMetGlySerLysT 1128  
10669 AACCTCGCGCAACCGCTCTCTGAGACGCGAGCGAGCGCTCTGCA 10718  
1128 hIlyseGluThrAspSerProSer..... 1135  
10719 CGAGGCGCATGCTCTCTCGAGAGTCAAGCGCGCGCATCCGCTGTCAGC 10768  
1136 ThrProGlyLysAspGlnArgSerMetIleLeuArgThrGlyLysT 1152  
10769 TCGCCCTCCGCGAGACTCGAGCGCTCAACCGCTGATCCCGCTGCTGTC 10818  
1152 hngIuGluIlePheHisSerLysArgArgArgProSerGluGlyArgLeu 1168  
10819 GG.....CTCATTCGAGAGAGAGAGCGCACTCG... 10849  
1169 ProAsnGly.....ArgAl 1173  
10850 CCGCGCTCTCTCTCTCTGAGAGCTCTCAACCGAGCGCGCTGCGCCTC 10899  
1173 ThrLysLysLeuLeuAspAsnSerHisLeuProAlaThrPheLysAla 1190  
10900 GAGCTGAGAGACTCTCTCGCGCGCTCAACCGCGCGAGAGCTCTCTCTC 10949  
1190 erSerSerProGluLysGlu.....GlyArgAlaSerGlnAla 1203  
10950 CACCTACCCCTTCAGAGAGAGAGGCTTGTGCTGAGCTCTCAAGAGAG 10999  
1204 ArgVal..... 1205  
11000 AACGCTCCGAGCTGCTCTCCGAGCGCTGAGCTCGCGCGACCAATCCG 11049  
1206 .....ProLysProGlyAla..... 1210  
11050 CTGCTGCGCGCGCGCTGCGCTGCGCGCGCGCGCTCTCTCTCTTAC 11099

1211 .....GlySerLysLeuSerAspArg 1217  
11100 AGGCGGCTCTCCCTCGAGAGACACCGTGGCTCGAAGGCCATGCCGTCT 11149  
1218 ProLeuHisAlaLeuLysArgLysSerAlaPhe..... 1228  
11150 TCGGCACACCATCTCTACGGGACCGGCTTCTCGAGCTCGCCCTGCAC 11199  
1229 .....MetAlaProValProThrLys.....LysArgA 1238  
11200 GTCCGCCACCGCTCGGCTCGACACCGTCTGAGAGCTCAGCTCGAGGC 11249  
1238 snLeuValLeuArgHisGlySerSerSerSerSerSerSerSerSerAlaMet 1254  
11250 CCTCTCGCTCTCCATCGCAGGACACCGCTCTCTCTCAGATCTCGCTCG 11299  
1255 Gly..... 1255  
11300 GCGCGTGGAGCAGCAGGAGGAGGCGCTCTCTTCCATAGCGGACAA 11349  
1256 .....GluMetG 1258  
11350 GAGGACGGCTTCAGGATGGCCCTGGACTCGCCAGCCAGCGGCTCTCT 11399  
1258 LysArgArgGlyLeuArgValProProProSerSerArgGlyCysLeu 1274  
11400 CTCGGCGGCGACCCCATCCCTCTCGCGGATCTCCACGAGTGC..... 11443  
1275 LeuProArg.....LysProSe 1280  
11444 CTCCTCGAGTGGCATCCCGGTGGACCTCGAGGCTCTAGCGAACCTC 11493  
1280 rPro.....ProArgAlaMetAlaSerLeuProG 1290  
11494 GCCAACCTCGGGCTTGCTACGCGCCGAGTCCAGGGCTCCGCTCCGT 11543  
1290 lnSerSer.....His 1293  
11544 CTCACAGCGCGGACAGAGCTTTTCGAGGACCAAGCTCCCGGAGCGG 11593  
1294 ProArgArg...ProProPheLeuProGlnAlaLeuSerSerAlaAlaPhe 1309  
11594 CCGAAAGAGTCCGCGCGGTGTGCCC...TCCACCTCGCTCGCTCGAC 11640  
1309 eGlnGlyAlaMetLysThrLysValLeuProProArgLysGlyArgGlyL 1326  
11641 AGCGCCCTTCATGCATGSCCTTTGAGGACGAGCAGAGGAGCGGTGC 11690  
1326 eu..... 1326  
11691 TCTGCCCTCTCTGGAGGAGTCTCGCTCGCTCGCTCGGTCGCCACCA 11740  
1327 .....LysLeuGluAlaIleValGlnLysIleThrSerProSer..... 1339  
11741 CTTTGGCGGTGGCTTCCACCGTCCCAAGGTGAATCTTCCCTCTCGATC 11790  
1340 .....LeuLysLysPheA 1344  
11791 GTCGTGGCGGACGCGCGAGTGAACCTCTTGCCTCGGTGCAAGCGTCTGC 11840  
1344 laCys..... 1345  
11841 CATCGGAGAGCTCCGCGCGGAGCTCCGACCCCGGAGCTTCCACC 11890  
1346 .....LysAlaPro..... 1348  
11891 ATGATGGCTCTTCCGGCTCGACTGGAGGAGCTCCAAAGCCCACTTCA 11940  
1348 ..... 1348  
11941 CGGCTCGCCCGGAGGCGGCTCTCTTCGCGACAGGCGGCGCAGATCT 11990  
1349 .....GlyAlaSerProGlyAsnProLeuSer...ProSerLeu 1360

11991 CGCGCTCGACGCCCGCTCGCCGCTAGCGGACCTCGCTGCCCTCCGAA 12040  
1361 SerAspLysAspArgGlyLeuLys...GlyAlaGlyGlySerProValG 1376  
12041 GCGCCTCGACAGGGGCTTGGCTCCCGGCTCGCTCGCTCGCC..... 12085  
1376 yValGluGluLeuValAsnValGlyThrGlyGlnLysLeuProThr. 1392  
12086 ....CCTTCATCGATCGACCGGACGCTCGTCCGAGCGCCACGA 12131  
1393 .....SerGlyAlaAspProLeuCysArg.....AsnProThrAsn 1404  
12132 GGCACCGCGCTCGCATCGCTCTTGAAGCTGCTCGCGGAGAC 12181  
1405 ArgSerLeuLysGlyLysLeuMetAsnSerLysLysLeuSerSerThrAs 1421  
12192 GCTCGCTCGCTCGGCTCGTCTCGTCCACCGAGCGCGCTCGCCACC 12231  
1421 pCysPheLysThrGluAlaPheThrSerProGluAlaLeuGlnProGlyG 1438  
12232 CACACCGAAGACGCTCAAGACCTCG.....CTCAGCGCGCTCTG 12275  
1408 LysThrAlaLeuAlaProLysArgSerArgLysGlyArgAlaGlyAla 1454  
12276 GGGCTCGCGCTCGCGCAAGAGTGAACCCAGACCTCCCGCTCTTCC 12325  
1455 HisGlyLeuSerLysGlyProLeuGluLysArgProTyrLeuGlyProAl 1471  
12366 TCGTCCACATCG.....ACCTCAGCGAGGCTCCACGAGCGCTG 12366  
1471 aLeuLeuLeuThrProArgAspArgAlaSerGlyThrGlnGlyAlaSerG 1488  
12367 CTAGCGCTCGACACAGGAGAACGCCAGCTCGCTCCGCAACGGGAA 12416  
1488 LuAspAsnSerGlyGly..... 1493  
12417 ACCCTCATCCGAGTGGGCGCAACACGCTCGACGAGCGCTCATCC 12466  
1494 .....GlyGlyLys.....LysProLysMetGluG 1502  
12487 CGCGCAGCAGCCCGCTGGCGCTCCATATTCGACAAAGCACCTTC 12516  
1502 uLeu...GlyProAlaSerGlnProProGluGlyArgProCysGlnProG 1518  
12517 GACGCTCGCTCGTCAAGCGCGG.....AGCGCAGCGCGCT 12560  
1518 lnThrArgAlaGlnLysGlnProGlyHisThrAsnTyrSerTyrSer 1534  
12561 CGCACGCGCAAGTCCGCTCGCTCGACGCGGCGGCTCACTTCC 12610  
1535 LysArgLys.....ArgLeuThrArgGlyArgAl 1544  
12611 GCGATGCTCGTGCACACCTTGGCTGATCCGCGGCGAGCGCGCGC... 12658  
1544 aLysAsnThrThrSerSerProCysLysGlyArgAlaLysArgArgG 1561  
12659 .....TCGAGCGCAAGCGCGGCGCTCTTACTGAAT 12692  
1561 lnGlnGlnValLeuProLeuAspPro..... 1569  
12693 CGGTCCAGGTCTCCGATACACCGTAGCGGCGGCTGGGTCT 12742  
1570 ...AlaGluProGluIleArgLeuLysTyrIleSerSerCysLysArgLe 1585  
12743 TCGCGCGAGCTTTGTC..... 12760  
1585 uArgSerAspSerArgThrProAlaPheSerProPheValArgValGluL 1602  
12761 .CCAGCGCATCGCGACCGCGCTGATCT..... 12790  
1602 ysArgAspAlaPheThrThrIleCysThrValValAsnSerProGlyAsp 1618





749 uasnleuGlYAspAlaCysProArgTrpGlyLeuHisProGlyLeuT 766  
 766 hrlYsGLYleuGLInGlyLysAlaSerAspGlyIleSerLYsGLY 782  
 9553 .....CGGCGCGGCTCTCCGCTTCGAGCTCTCCGACAC 9588  
 783 AsPThrHis.....GluAlaSerAlaCysLeuGlyPheGlnG 795  
 9589 AACGCCACGTCATCTCGAGAGAGGCTCCGCGCGCTG.....TC 9629  
 795 uGlnAspProProGlyGlyLysValAlaSerLeuProGlyAspPheLYsG 812  
 9630 GGGCGAGCCCGCACCTCCACAGCGGCTCGACCGCTCCCGCGCGCT 9679  
 812 InGluGluValGlyGlyValLYsGlnGluAlaGlyGlyLeuLeuGlnCys 828  
 9680 GTCCCGTGTCTCTGTGCGCAGAGAGCGCGCGCTCCGCGCGCAGGCG 9729  
 829 ProGlu.....ValAlaLYsAlaAspArgTrpLeuGlnAs 840  
 9730 AAGCGGCTCCGCGACACCACTCTCTCGCCACAGACGACCTCGCTTTCGA 9779  
 840 pSerArgHisCysCysSerThr.....AlaAspPhe.GlyAspLeuPro 854  
 9780 TGTGGCTATTTCGACAGCCACCAACCGCGCCACCTTCAGACACCGCGCG 9829  
 855 LeuLeuProProThrSerArgLYsGlyAspLeuGluAlaGlnGlnTyr 871  
 9830 CTCTCTGTGCGCGCGACCGCGACG..... 9853  
 871 rSerSerLeuCysGlnLeuLeuGlySerProGlnGlnArgProGlyMetC 888  
 9854 .AcTCTCTCTCGCGCTCTCGCTCGCGCGACAGACAAAGCCG..... 9895  
 888 InAspProLeuSerProLYsAlaProLeuIleCysThrLYsGlnGluVal 904  
 9896 .....CCCGAGACCGCTTCGCGCGCGAGCGAGAAAGCGACG 9931  
 905 GlnGluValLeuAspSer.....LYsAla.....GlyTr 914  
 9932 GCAGAGTGTCTCTGCTTCCTCGGCGACAGCTCGCAGTGGAGAGGATG 9981  
 914 pGlySerProCysHisLeuSerGlyGlnSerValIleLeuLeuGlyProT 931  
 9982 GCGCTCTCTCGCTCGACCTCTCGCGCGCTCTCGCGCTCAGCTCGAAGC 10031  
 931 hValGlyThrGlnSerLYsValGlnSerTrpPheGlnSerSerLeuSer 947  
 10032 ATGCGAGCGCGCTCTCGCTCTCGACGTCG..... 10060  
 948 HisMetLYsProGlyGlnGlnGlyProAsp.....GlyGluArgAlaPro 962  
 10061 .....AGTGAACCTGCTCTCGCTCTCGCGCGCGAGAGGAGCGC 10101  
 962 oGlyAspSerThrThrSerAspAla..... 970  
 10102 CCTCTCTCGACGCGCTCGACGTGTACAGCCGCTCTTTCGCTTCAT 10151  
 971 .....SerLeuAlaGlnLYsProAlaHisProAlaVal 981  
 10152 GGTCTCTCGCGCGCGCTCTGCGCTGCGCTGCGCGCGCGCGCGCG 10201  
 982 ProGluAlaProIleAlaLYsGlnGluProValProArgGlyLYsSerLe 998  
 10202 TCGTGGCGCACAGCCAGGCGAGATCGCGCGCTTCGTGCGAGCGCT 10251  
 998 uArgSerArgArgValHisArgGlyLeuProGlnGluAlaGlnAspSerProC 1015  
 10252 CTCTCCCTCGAGAGCGCGCGCGCTCGCG.....CCCT 10286  
 1015 ys.....ArgAlaProValLeuProLYsAspLeuLeuProGlu 1028

10287 GCGAGGAAAGCGCTCACACCGCTCGGCGCAAGGCGGAGGCGCGCG 10336  
 1029 SerCysThrGlyProProGlnGlyGln.....MetGlnLYsAla.. 1041  
 10337 TCGAGCGTGGCGGCTCTCGACCTCCAGACCTTACCTGCTCTCCGAGGCGAC 10386  
 1042 .....GlyAlaProGlyArgGlyAlaSerGln 1051  
 10387 AGGCTCTCCACCGCGCGCTCCAGACCGCCAGGCGCTACCTGTATCCG 10436  
 1051 LYsLeuProArgMetCysThrArgSerLeuThrAlaLeuSerGlnProArg 1067  
 10437 CGAGCGCGCGCGCTCGAGCGGCTGCTCGACGCTCTACCGCGCCAGAG 10486  
 1068 ...ThrProGlyProProGlyLeuLeuThrThrProAlaPro..... 1080  
 10487 TGTTCGCGCGCAAGATCCGCTGTAGCTTACGCTCCACCTCGCGCCAGATG 10536  
 1061 .....ProAspLYsLeuGlyG 1086  
 10537 GAAGCGCTCCAGACGAGCTCGCGAGGCTTACGCAACATCTCTCTCG 10586  
 1086 LYsGlnArgAlaAlaPheLYsSerGlyLYsArgValGlyLYsProSer 1102  
 10587 GACGTGGAGCTCTCTCTTATT.....CGACCGTCA 10618  
 1103 ProLYsAlaAlaSer...SerProSerAspProAla..... 1113  
 10619 CCGGACACGAGCTCGACGCTCCAGCTCGACGCGGCTACTGTATCGA 10668  
 1114 .....AlaLeuProValAlaSerAspSerProMetGlySerLYsT 1128  
 10669 AACCTCCGCGCAACGCTCTCTGTGTGAGCGGACCGAGCGGCTCTCGA 10718  
 1128 hLYsGlnThrAspSerProSer..... 1135  
 10719 CATGTGGATGCTCTCTCGGCGAGGTACAGCCCGACCGCTCTCGACG 10768  
 1136 ThrProGly LYsAspGlnArgSerMetIleLeuArgSerArgTrpLYsT 1152  
 10769 TCGGCTCTCGCGGAGCTCGAGGCGCTCACCGCTGATCCGCTGCTGTC 10818  
 1152 hGlnGlnIlePheHisSerLYsArgArgArgProSerGlnGlyArgLeu 1168  
 10819 GG.....CTCATTCGACAGAGAGAGGCGACCTCG..... 10849  
 1169 ProAsnCys.....ArgAl 1173  
 10850 CCGGCTGCTCTCTCTCGGCGGAGGCTCTTACCGAGGCGCTCGGCGTC 10899  
 1173 ArgLYsLYsLeuLeuAspAsnSerHisLeuProAlaThrPheLYsValS 1190  
 10900 GACTGGAGAGACTTCTTCGCGCTTACGCTCCCGCAGAGGCTCTCTCC 10949  
 1190 SerSerSerProGlnLYsGln.....GlyArgValSerGlnArgAla 1203  
 10950 CACCTACCGCTTCAGCAGAGAGCGGCTTGTGCTGTGAGCTCTCACAGAGC 10999  
 1204 ArgVal..... 1205  
 11000 AACGCTTCGACGCTGCTCTCGACAGGCTGACCTGAGCGGCGACCATCCG 11049  
 1206 .....ProLYsProGlyVala..... 1210  
 11050 CTGTCTGGCGCGCGCTGCTTGGCGACCGCGCTGTCTTTCAC 11099  
 1213 .....GlySerLYsLeuSerAspArg 1217  
 11100 AAGGCGGCTCTCTCTCGAGAGCACCGGCTCTGAGAGCGCATGCGCTCT 11149  
 1218 ProLeuHisAlaLeuLYsArgLYsSerAlaPhe..... 1228

11150 TCGGCACACCATCTACCGGACCGGCTTTCGAGCTCGCCCTGCAC 11199  
1229 .....MetalaprovAlProThrLys.....LysArgA 1238  
11200 GTCGCCACCGCTCGGCTCGACCGCTGCAAGAGCTCAGCTCGAGC 11249  
1238 snLeuValLeuArgHisGlySerSerSerSerSerSerSerSerSer 1254  
11250 CCCTCGCTCCCATCGAGGACACGGCTCTCTCCAGATCTCCGTGG 11299  
1255 Gly..... 1255  
11300 GCGCGTGGACGACGACGAGGAGGCGCTCTCTTCATAGCCGACAA 11349  
1256 .....GluMetG 1258  
11350 GAGGACGGCTTCAGGATGGCCCTGGACTCGCCACGCGGCGCTCTCT 11399  
1258 LysArgArgGlyLeuArgValProProSerSerArgGlyCysLeu 1274  
11400 CTCGCCGCGACCCCATCCCTCTCGCGGATCTCCAGAGTGGC..... 11443  
1275 LeuProArg.....LysProse 1280  
11444 CTCCTCGAGTGCATCCGCTGGACCTCGAAGGCTCTACGCAACCTC 11493  
1280 rPro.....ProArgAlaMetAlaSerLeuProG 1290  
11494 GCCAAGCTCGGCTTGCTACGCGCCCGAGTTCAGAGGCTCTCGCTCG 11543  
1290 InSerSer.....His 1293  
11544 CTACAGCGCGGACGAGCTCTTTGCGAAGCAACCTCGCGAAGCGG 11593  
1294 ProArgArg.....ProProPheLeuProGlnAlaLysSerAlaAlaph 1309  
11594 CCGAAGAGTACGCGCGCTTGGCC.....TCCACCTGCTGCTCGAC 11640  
1309 eGlnGlyAlaMetLysThrLysValLeuProProArgLysGlyArgGlyL 1326  
11641 AGCGCCCTGATGCTGCTGCTTGGAGTACGAGGAGGAGGAGCTGCG 11690  
1326 eu..... 1336  
11691 TCTGCCCTTCTGTGAGCGGAGTCTGCTGCTGCTGCTGCTGCTGCTG 11740  
1327 .....LysLeuGlnAlaIleValGlnLysIleThrSerProser..... 1339  
11741 CTTGCGCTGCTGCTTCCACCGTCCCAAGGTAATCTCTCGTCTCGATC 11790  
1340 .....LeuLysIysPheA 1344  
11791 GTCTGCGGACGCGCGAGTACCTCTTGTGCTGCTGCTGCTGCTGCTG 11840  
1344 laCys..... 1345  
11841 CATGCGGACGAGCTCGCGCGGAGTCCGACCCCGGAGCTTCCGACC 11890  
1346 .....LysAlaPro..... 1348  
11891 ATGATGCGCTCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11940  
1348 ..... 1348  
11941 CCGCTGCGCGCGGACGCGCTCTTCTGCGACAGGCGGCGGCTGCT 11990  
1349 .....GlyAlaSerProGlyAsnProLeuSer...ProSerLeu 1360  
11991 CGGCTGCGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 12040  
1361 SerAspLysAspArgGlyLeuLys...GlyAlaGlyLysProValG 1376  
12041 GCGCCCTGACGAGGCGCTTCCCTCCCGGCTGCTGCTGCTGCTGCTG 12085





[illegible]



1128 hrLysGluThrAspSerProSer..... 1135  
1136 ThrProGlyLysAspGlnArgSerMetIleLeuArgSerArgThrLysT 1152  
1139 eGlnGlyAlaMetLysThrLysValLeuProArgLysGlyArgGlyL 1326  
1326 eu..... 1326  
1327 .....LysLeuGluAlaIleValGlnLysIleThrSerProSer..... 1339  
1344 .....LysLysPheA 1344  
1348 .....LysAlaPro..... 1348  
1349 .....GlyAlaSerProGlyAsnProLeuSer...ProSerLeu 1360  
1361 SerAspLysAspArgGlyLeuLys...GlyAlaGlyGlySerProValG 1376  
1376 yValGluGluGlyLeuValAsnValGlyThrGlyGlnLysLeuProThr. 1392  
1393 .....SerGlyAlaAspProLeuCysArg.....AsnProThrAsn 1404  
1405 ArgSerLeuLysGlyLysLeuMetAsnSerLysLysLeuSerSerThrAs 1421  
1421 pCysPheLysThrGluAlaPheThrSerProGluAlaLeuGlnProGlyG 1438  
1438 lyThraLeuAlaLeuProLysLysArgSerArgLysGlyArgAlaGlyAla 1454  
1454 HisGlyLeuSerLysGlyProLeuGluLysArgProFyrLeuGlyProAl 1471  
1471 aLeuLeuLeuThrProArgAspArgAlaSerGlyThrGlnGlyAlaSerG 1488  
1488 luAspAsnSerGlyGly..... 1493  
1494 .....GlyGlyLys.....LysProLysMetGluG 1502

27897 CGAAAAGAGATGCCCGCGGTTCGCC...TCCACCGCTGCGCTCGAC 27943  
1309 eGlnGlyAlaMetLysThrLysValLeuProArgLysGlyArgGlyL 1326  
27944 AGGCGCTGCATGCACTGCGCTTTGAGGAGGAGCAGAGAGGAGCGTGC 27993  
1326 eu..... 1326  
27994 TGTGCGCTTCGTGGAGGAGTCTGCTGCGCTCCGCTGCGTCAACCA 28043  
1327 .....LysLeuGluAlaIleValGlnLysIleThrSerProSer..... 1339  
28044 CTTGCGCGTGCCTTCCACGCTCCCAAGGTGAATCCTCGTCTGATC 28093  
1340 .....LeuLysLysPheA 1344  
28094 GTCTGGCGGAGCGCGCAGGTGACCTTGTGCTCGGTGCAAGGCTGC 28143  
1344.....LaCys..... 1345  
28144 CATGCGGAGAGCTCGCGCGAGCTCCGACCGCGCAGCTTCCACCC 28193  
1346 .....LysAlaPro..... 1348  
28194 ATGATGCGCTTCTCCGCTGCTGAGTGGAGGAGCTCCAAAGCGCCACTCA 28243  
1348..... 1348  
28244 CGCGCTGCGCGCGCGAGCGGTCTTCTCGGACAGCGCGCCACGATCT 28293  
1349 .....GlyAlaSerProGlyAsnProLeuSer...ProSerLeu 1360  
28294 CGCGCTGAGCGCGCTCGCGCGCTGAGCGCGCTCGCTCGCTCGAA 28343  
1361 SerAspLysAspArgGlyLeuLys...GlyAlaGlyGlySerProValG 1376  
28344 GCGCTTGCACAGCGCGCTTGTGCTTCCGCGCTGCTGCTGCTGCTG 28388  
1376 yValGluGluGlyLeuValAsnValGlyThrGlyGlnLysLeuProThr. 1392  
28389 .....CCTTCATCATGAGCGCGCGAGGAGCTCGTCCGCGCGCGCTCG 28434  
1393 .....SerGlyAlaAspProLeuCysArg.....AsnProThrAsn 1404  
28435 GCGCACCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 28484  
1405 ArgSerLeuLysGlyLysLeuMetAsnSerLysLysLeuSerSerThrAs 1421  
28485 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 28534  
1421 pCysPheLysThrGluAlaPheThrSerProGluAlaLeuGlnProGlyG 1438  
28535 CACACCGAAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 28578  
1438 lyThraLeuAlaLeuProLysLysArgSerArgLysGlyArgAlaGlyAla 1454  
28579 GGGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCTG 28628  
1455 HisGlyLeuSerLysGlyProLeuGluLysArgProFyrLeuGlyProAl 1471  
28629 TGTGCGATCG.....ACCTCAGCGAGGCTTCCCGCGCGCTGCTGCTG 28669  
1471 aLeuLeuLeuThrProArgAspArgAlaSerGlyThrGlnGlyAlaSerG 1488  
28670 CTAGCGCGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 28719  
1488 luAspAsnSerGlyGly..... 1493  
28720 ACCCTCATCCCGAGGTTCGCGCAACCGCTGCGAGCGCGCTCATCC 28769  
1494 .....GlyGlyLys.....LysProLysMetGluG 1502

[illegible]

257 CAGCAGCGACTGCCGCTGCCAGGCTGTGGCAGGGGGCCCACTCTCTAY 306  
244 ..... 244  
307 AAGAGYTCACAGCACCRCCTGCCAGCCTCATGATAGCGCGAGTATGC 356  
245 ..... GlnProHisAspArgp 250  
357 CAATGGGAACCTGGCTCCAGGCAACGGGTCCAGCAGCAGCCCATGACAGGC 406  
250 roleuThraLaserSerSerLeuAlaProGlyGlnArgValGln. .... 264  
407 CGCTGACTGCCAGCTCCAGCTGCGCCCGGGGCGAGCGGTCCAGGGGCA 456  
264 ..... 264  
457 GCGGGTCCAGCAGCCYCATGAYAGCCGCTGAGTGCARCYKCSARCTGG 506  
264 ..... 264  
507 CYCRRGGGCAACGGGTCCAGATCTTACAGCTTACAGCCTGGCGGCTT 556  
265 ..... AsnLeuHisAlaTyrGlnSerGlyArgLeuSerTyrAspGln 278  
557 GGCTAGCAATCTTCATGCTACAGCTGCGGGCGGCTGAGCTATGACAG 606  
279 GlnGlnGlnGlnGln. Gln. .... 284  
607 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 656  
284 ..... 284  
657 GCCTYRGCTAYGACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 706  
284 ..... 284  
707 ACTTACAGCGCTCCACACACCCAGGAACACTCCACAGCAGCAGCAGC 756  
285 ..... Gln 285  
757 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 806  
286 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 302  
807 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 856  
302 rLeuHisTyrGlnAsnLeuAlaLysTyrGlnHisTyrGlyGlnGlnGly 319  
857 MCTCCATTACCAAGACCTCGCAAGTACCAACACTATGGACAGCAGGCC 906  
319 InGlyTyrCysGlnProAspAlaAlaValArgThrProGlnGlnTyr 335  
907 AGGGCTACTCTCCACTACCAAAACC. .... 932  
336 GlnThrPheSerProSerSerHisSerProAlaArgSerValGlyAr 352  
932 ..... 932  
352 gSerProSerTyrSerSerThrProSerProLeuMetProAsnLeuGlu 369  
933 .TCGCCAAGTATCAGCACTACGCGCAGCAAGCAGCAGGCTACTGCCAGCC 981  
369 snPheProTyrSerGlnGln 375  
982 AGCAC...TACGGGCGACAA 998  
seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NAL1986.DAT.AAT06769  
seq\_documentation\_block:  
ID AAT06769 standard; DNA; 28598 BP.  
XX  
AC  
AT06769;  
XX

15-OCT-1996 (first entry)  
XX  
XX Sorangium cellulosum soraphen gene cluster.  
XX  
XX Antipathogenic substance; soraphen; phenylpyrrole; antibiotic;  
XX  
XX fungicide; pesticide; myxobacterium; ss.  
XX  
XX Sorangium cellulosum.  
XX  
XX WO9533818-A2.  
XX  
XX 1-DEC-1995.  
XX  
XX 30-MAY-1995; 95WO-IB00414.  
XX  
XX 08-JUN-1994; 94US-0258261.  
XX  
XX (CIBA ) CIBA GEIGY AG.  
XX  
XX Beck JJ, Gaffney TD, Hammer PE, Hill DS, Lam ST;  
XX  
XX Ligon J, Ryals JA, Schupp T, Uknes SU;  
XX  
XX WPI; 1996-040226/04.  
XX  
XX New genes for biosynthesis of anti-pathogenic substances - pref.  
XX  
XX Pyrrolnitrin and soraphen, useful for disease control in plants  
XX  
XX Claim 3; Page 140-158; 190pp; English.  
XX  
XX This is the soraphen gene cluster from S. cellulosum. The gene  
XX  
XX cluster may be expressed recombinantly to produce soraphen, or  
XX  
XX expressed in a transgenic plant for disease-resistance.  
XX  
XX Sequence 28598 BP; 4031 A; 11703 C; 8263 G; 4601 T; 0 other;  
SQ  
alignment\_scores:  
Quality: 487.00 Length: 2135  
Ratio: 0.553 Gaps: 121  
Percent Similarity: 41.265 Percent Identity: 22.670  
alignment\_block:  
US-09-508-821B-6 x AAT06769 ..  
Align\_seg 1/1 to: AAT06769 from: 1 to: 28598  
18 GlnGlnThrSerGlnGlnThrSerArgLeuGlu.....As 29  
22787 CAGGAACTCCAGACCTGCTGCTGCTCGACCCCTCGACGCTTGTGTC 22836  
229 nTyrArgGlnProSerGlnAlaGlyLeuSerCysAspArgGlnArgLeuL 46  
22837 CTTTCTGCTCCATCGCAGCGCTCTGCGGCGCGGCAACAG..... 22879  
46 euAlaLysAspTyrTyrAsnProGlnProTyrProSerTyrGlnGly 62  
22880 ..... CCGATACGCGCGCG.....GAAACGCT 22903  
63 AlaGlyThrProSerGlyThrAlaAlaValAlaAlaAspLysTyrHi 79  
22904 TCCTCGAGCGCTTGGCGACCGAGCGCGCTTGGACAGCGCGACACG 22953  
179 s.....ArgGlySerLysAlaLeuProThrGlnGlnGlyL 91  
22954 TCGTGTGTGGGCGCGGTGGGCGCGGTGTATATTATTCACGGGCGC. 23002  
91 euGlnGlyArgProAlaPheProGlyTyrGlyValGlnAspSerSerPro 107  
23003 .....CCCTGGCAGCCGAGCTGGACACACGCTGCTGTGTCGCGCA 23041  
108 TyrProGlyArgTyrAlaGlyGluGluSerLeuGlnAlaTyrPglyAlaPr 124  
23042 TGCCCCCTTCGCTGCGGTGGGCGCTCGCGCAAGCCCTGGAGCAGCAGC 23091



24437 .....AGTCTCTCCGGCGCGCTCCACACCCCGCATCACCC 24475  
668 ProAspSerLeuGlnLeuAspGlyGlyValLeuLeuLeuGlyProThrValGI 933  
24476 CGCCGCGG.....TCGACGTCGTCGAGTCGATGGCAGCGGACCC 24516  
684 oglyLeuPheLeuAspProSerValAlaPheAlaThrProAspProLysL 701  
24517 T...CGCTGGGAGACCCATCGAGTCCAAAGCCCTCCCTCTCGCGCGCTCAAGACCA 24563  
701 ysThrThrGlyProLeu.....SerPheGlyThrLysPro 712  
24564 CGACGCGACACCGCTGAAAGCCCTCTCTCTCGCGCGCTCAAGACCA 24613  
713 ThrLeuGlyValProAlaPro..... 719  
24614 ACATCGCCCATCTCGAGCGCGCTCCGCGCTCCGCGGCGTCCGCAAGATG 24663  
720 AspProThrThrAlaAlaPheAspCysPheProAspThrThr..... 733  
24664 GTGGCTGCTCCGCGACACCGCTCCGCGCGCTCCGCGGCGTCCGATA 24713  
734 .....AlaAlaSerSerAlaAspSerAla..A 742  
24714 ACGCAATCCCTCATCGAGTGGGAGCGCTCGCCATCGACGTCGTCGATA 24763  
742 snProPheAlaThrProGluGlnLeuGlyAspAlaCysProArgTrp 758  
24764 CCGCGAGCGCTTGGCCCGCCGACAA.....GATGGCAGTCCCGC... 24804  
759 GlyLeuHisProGlyGluLeuThrLysGlyLeuGluGlnGlyLysAl 775  
24805 .....CGCGCGGCGATCTC 24818  
775 aserAspGlyLeuSerLysGlyAspThrHis.....GluAla 786  
24819 CGCTTCGATCTCGGCGACCAACCGCGCTCATCTCGAAGAGGCTC 24868  
788 erAlaCysLeuGlyPheGlnGluAspProGlyGluLysValAla 804  
24869 CGCGCGCGCTG.....CGCGCGGCGATCTC 24897  
805 SerLeuProGlyAspPheLysGlnGluValGlyGlyVal..... 818  
24898 TCACAGCGCGCTCGCAAGCGCTCCGCGGCGTCCGCGCTCTCTCTGTC 24947  
819 .....LysGluGluAlaGlyLeuLeuGlnCysProGlu..... 830  
24948 GCCCAGGCGAGCGCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 24997  
831 .....ValAlaLysAlaAspArgTrpLeuGluAspSerArgHisCysCys 845  
24998 ACCTCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 25047  
846 SerThr.....AlaAspPhe.GlyAspLeuProLeuLeuProThrS 860  
25048 GCCCAGCG 25097  
860 erArgLysGluAspLeuGluAlaGluGlnLysSerSerLeuCysGlu 876  
25098 CGCGCAG.....AGCTCTCTCCGCGC 25120  
877 LeuLeuGlySerProGluGlnArgProGlyMetGlnAspProLeuSerPr 893  
25121 TCGACTCTGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 25149  
893 oLysAlaProLeuLeuLysThrLysGluGlnValGluGlnValLeuAspS 910  
25150 CCGAGCAGCG.....TCCTCGGCGGAGGAGGAGCGCGCGCGCGCGCG 25190  
910 erLysAla.....GlyTrpLysSer 916

25191 CFTCT 25240  
917 ProCysHisLeuSerGlyGlyLeuSerValLeuLeuLeuGlyProThrValGI 933  
25241 CGTGTCT 25290  
933 yThrGluSerLysValGlnSerTrpPheGluSerSerLeuSerHisMetL 950  
25291 CGCGCGCT 25312  
950 ysProGlyGluGlyProAsp.....GlyGluArgAlaProGlyAsp 964  
25313 .AGTGGACCT 25360  
965 SerThrThrSerAspAla..... 970  
25361 TCGACCGCT 25410  
971 .....SerLeuAlaGlnLysProAsnLysProAlaValProGluA 984  
25411 CTGGCGCGCT 25460  
984 LaProLeuAlaLysLysGlnProValProArgGlyLysSerLeuArgSer 1000  
25461 CCACGTCAGGCGAGATCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 25510  
1001 ArgArgValHisArgGlyLeuProGluGlnGluGluAspSerProCys..... 1015  
25511 TCGAGGACGCGCGCGCATCCCG.....CCCTGCCGACG 25545  
1016 .....ArgAlaProValLeuProLysAspLeuLeuProGluSerCysT 1031  
25545 AAAGCGCTCACACCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 25595  
1031 hrGlyProGlnGlyGln.....MetGluGlyAla...GlyAla 1043  
25596 CGCGCGCT 25645  
1044 ProGlyArgGlyAlaSerGluGlyLeuProArgMetCys..... 1056  
25646 CCATCGCGCGCTCAACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 25695  
1057 .....ThrArgSerLeuThrAlaLeuSerGluP 1066  
25696 GCGCGCATCGACGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 25745  
1066 roArg.....ThrProGlyProProGlyLeuThr... 1075  
25746 CCGAAAGTCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 25795  
1076 ...ThrThrProAlaPro.....ProAspLysLeuGlyLysGI 1088  
25796 TCCAAAGAGCTCGCGCGAGTCTAGCAACATCTCTCTCTCTCTCTCTCTCT 25845  
1088 nArgAlaAlaPheLysSerGlyLysArgValGlyLysProSerProLysA 1105  
25846 GAGCT 25877  
1105 laAlaSer...SerProSerAsnProAla..... 1113  
25878 CAGGCTCGACGCT 25927  
1114 AlaLeuProValAlaSerAspSerProMetGlySerLysThrLysGI 1130  
25928 GGCAACCGCT 25977  
1130 uThrAspSerProSer.....ThrProG 1138  
25978 CATCGCT 26027  
1138 lyLysAspGlnArgSerMetLeuLeuArgSerArgThrLysThrGlnGI 1154  
26028 CCGCGAGACCTCGCAGCGCTCACCGCTCTCTCTCTCTCTCTCTCTCTCTCT 26072

```

1154 uIlePheHisSerLySArgArgProSerGlu..... 1165
      :::::::::::::::::::::
26073 ....CTCCATTGACGGCCAGCAGAGGCCAACCTCCGCCGCTCCCTCC 26118
      :::::::::::::::::::::
1166 .....GlyAlaGlu...ProAsnCysArgAlaThrLysLysLeuAsp 1179
      ||| ||| ||| ||| ||| ||| ||| ||| |||
26119 TGGAGCGAGCTCTACACCGGAGGCTCGGCTGCATCGAGAACGCTTC 26168
      :::::::::::::::::::::
1180 AsnSer.....HisLeuProAlaThrPh 1187
      :::::::::::::::::::::
26169 GCGGCCCTTGCTCCCGCAGAGCTCCCTCCGACCTACCCCTCCAC 26218
      :::::::::::::::::::::
1187 euLysValSerSerSerProGlnLysGlu...ArgValSerGlnArg 1203
      :::::::::::::::::::::
26219 GCGAGAGCTTGTGGCTCGACCGCTCCACAGAGAGGCTCTCCCTCGAG 26268
      :::::::::::::::::::::
1203 LalaArgProLysProGly.....AlaGlySerLysLeuSerAspArg 1217
      ||| ||| ||| ||| ||| ||| ||| ||| |||
26269 TCCGACAGCTGACCTCGGCGCAGACCCGCTGCTCGGCGCGCGTGC 26318
      :::::::::::::::::::::
1218 Pro.....LeuHis.....AlaLeuLysArgLy 1225
      ||| ||| ||| ||| ||| ||| ||| |||
26319 CCTCCGCCACCGCGAGAGGCTTTGCTTCACAGAGAGGCTCTCCCTCGAG 26368
      :::::::::::::::::::::
1225 sSerAlaPheMetAlaProValProThrLysLysArgAsnLeuValLeu 1242
      :::::::::::::::::::::
26369 A.....GCACCCGTTGCGTCGACAGACACAGCTGT.....CTTC 26400
      :::::::::::::::::::::
1242 rGHisGlySerSerSerSerSerAsnLaserAlaMetGlyGluMetGly 1258
      ||| ||| ||| ||| ||| ||| ||| ||| |||
26401 GGGATACCCCTGTC..... 26414
      :::::::::::::::::::::
1259 ArgArgArgGlyLeuArgValProProProSerSerArg...GlyCysLe 1274
      :::::::::::::::::::::
26415 ....TCCAGAGGCGCGCTCGACGCTCCGACGCTCCCTGCARGTGC 26452
      :::::::::::::::::::::
1274 uLeuProArgLysProSerProProArgAlaMetLaserLeuProGlns 1291
      ||| ||| ||| ||| ||| ||| ||| ||| |||
26453 CCCATCTCGTGGCTCGACACCTCGAGAACGTCACGCTGCAGCCCCC 26502
      :::::::::::::::::::::
1291 erSerHisPro.....ArgArgProProPheLeuProGlnAla 1303
      ||| ||| ||| ||| ||| ||| ||| ||| |||
26503 CTCGCTCTCCATCGACGAGGCGCGCTCTCTCCAAATCTCCGCGGCG 26552
      :::::::::::::::::::::
1304 ArgLeuSerAlaAlaPheGlnGlyAlaMetCysThrLysValLeuPro 1320
      ||| ||| ||| ||| ||| ||| ||| ||| |||
26555 GCGCAGAGGCTGTGAGACAGAGGCGCTCTCCGTTCAATACG...GGGCG 26599
      :::::::::::::::::::::
1320 oArgLysGlyArgGlyLeuLysLeuGlnAlaIleValGlnLysIleThrS 1337
      ||| ||| ||| ||| ||| ||| ||| ||| |||
26600 ACGACGCGGCTTCAGAGAGGCGCTCGATCGACACCGACGCGCTCTC 26649
      :::::::::::::::::::::
1337 er.....ProSerLeu..... 1340
      ||| ||| ||| ||| ||| ||| ||| ||| |||
26650 GCGCAAGCTAAGCCGCTCCATTCTCTCGAGGCTCGCGAATGCCCCC 26659
      :::::::::::::::::::::
1341 .....Ly 1341
      :::::::::::::::::::::
26700 CTCGAGCGCCACCCAGCTGACACCCAGGTTTCTACAGAGCCTCGAG 26749
      :::::::::::::::::::::
1341 sLysPheAlaCys.....LysAlaProGlyLysAspProGlyAsn. 1354
      :::::::::::::::::::::
26750 GCGCTGGGCTTGTATTAGGCCCGGAATTCAGGCGCTCCGCGCGCTCA 26799
      :::::::::::::::::::::
1354 ..... 1354
      :::::::::::::::::::::
26800 CAAGCGCGCGAGAGCTTTCGCGAAGCCAGCTCCGAGCGCGCG 26849
      :::::::::::::::::::::
1355 .....ProLeuSerProSerLeuSer..... 1361
      ||| ||| ||| ||| ||| ||| ||| ||| |||
26850 AAGAGAGCGCCGCTGTTTCCCTCCACACCGCGCGCTCGACAGAGCGC 26899
      :::::::::::::::::::::

```

1362 ..... AspLysAspArgGlyLeu 1368  
26900 TTGCAGAGCGCTCGCCTTTGTAGACGACGAGGAAAGGCTCTCGAGGATGCC 26949  
1368 ysgLAlaAgly..... SerProValGlyVal..... 1377  
26950 CTTCTCGAGGAGGAGATGCTGGCTCGCTCGAGTCGAGCACACCGTGG 26999  
1378 ..... GluGlyLeuValAsn\_ValGlyThrGly 1388  
27000 CGCGTGCCTTCCACCGCTCGAGGAGGAGATCGTGGCGTGCCTCT 27049  
1388 IndLysLeuProThrSerGlyAlaAspProLeuCysArg..... 1400  
27050 CGCGACGCCGACGAGGAGACCCATCGCTCGGTGCAAGCGCTCGGCATGC 27099  
1401 ..... AsnProThrAsnArgSerLeuGlyLysLysLeuMetAsnSerLys 1415  
27100 GCGCCGCGCTCGCGAGGACGCTCGGACACCGGAGGAGCGTCCACCTCGA 27149  
1415 sLysLeuSerSerThrAspCysPheLysThrGluAlaPheThrSerProG 1432  
27150 TGCCCTCTTCGCGATGATGAGGAGGAGCGTGAAGAACCCGACGACG 27198  
1432 IuAlaLeuGlnProGlyGlyThrAlaLeuAlaProLysLysArgSerArg 1448  
27199 .. CCGATGCCCGCCGAGGCGGTGCCCTCC..... 27225  
1449 LysGlyArgAlaGlyAlaHisGlyLeuSerLysGly..... ProLe 1462  
27226 ..... GCGACAGAGATGTCGACCTCGGAGACCAAGGCTCT 27263  
1462 uGluLysArg\_ProGlyLeuGlyProAlaLeuLeuThrProArgAsp 1478  
27264 CGACCGCTTACCGACCTTG.... CTGCTTCACGACGCGCCCTCGAC 27307  
1479 ArgAlaSerGlyThrGlnGlyAlaSerGluAspAsnSerGlyGlyGly 1495  
27308 AGGCGGCTTGCGCTCGACAGGCTGTCATCGCCCTCA..... TC 27348  
1495 yLysLysProLysMetGluLeuLeuGlyProAlaSerGlnProGlnG 1512  
27349 GCTCTGCCGAGGAGGACATCATCGAGCGCCCGGAGAACCGACGCGCA 27398  
1512 LysArg..... ProCysGlnProGlnThrArgAlaGlnLysGlnProGly 1526  
27399 CGCGCTCGCCTCTTCGACAGCGTGGCTCGCGAGACCGCTCGCGCTCT 27448  
1527 HisThrAsnLysSer..... Tyr 1533  
27449 CGCGCTCGCCTCTCTCAACCGAGCGGCGCGTGGCGACCAACCGCTGAAGA 27498  
1533 rSerLysArgLysArgLeuThrArg..... Gly Arg 1543  
27499 GACGTCAAGGCGCTCGCTCAAGCGCGCTCTGCGGAGTCTGCTCGCTCGC 27548  
1544 AlaLysAsnThrThrSerSerProCysLysGlyArgAlaArgArg 1560  
27549 CGAGAGCGACACCGACGAGCC.... TCTGTGCTGTGCACTCG 27592  
1560 gGlnGluGluValLeuProLysAspProAlaGlnProGluLeuArgLeu 1577  
27593 AGGACAGCGAGGCTTCGACGAGCGCTCTGCGGCGCGTCAAGCC... 27638  
1577 ysiThrLysSerCysLysArgLeuArgSer..... 1587  
27639 ..... AAGAGACGACGAGATCGGCTCCGCAAGGCA 27671  
1588 ..... AspSerArgThr 1591  
27672 ACCCGCTTAAGCAAGCTCTCATACGAGGAGGACCGGACGCTCGCG 27721  
1591 rProAlaPheSerProPheValArgValGluLysArgAspAlaPheThr 1608  
b1



```
27722 CCTGGTGGCGCCGCCCTCGCTGCTAAACACGAGC..... 27758
1608 hrIleCysThrValAlaSerProGlyAspAlaProIleProHisArg 1624
27759 .....CAAGCACCCTGCTCCACCTC 27779
1625 LysPro.SerSerSerAlaSerSerSerSerSerSerPheSerL 1641
27780 GCGCAGGCGCGGAGCGCTCGGGTGTCTGATGCTTGGAGC...GAGC 27826
1641 euAspAlaIleValSerLeuAlaThrLeuProGlyGlySerIleLeu 1657
27827 TGAAGCTCTGGGGCTTGGCTC...ACCTCGCGCGGTGGAGCTGGCC 27873
1658 GlnProArg.....ProSerLeu...Pr 1664
27874 GATCCACGCGCTCTAAAGGACCTCTGTGATAACATTCGAGCGCTCACCC 27923
1664 oLeuSerSerThrMetHisLeuGlyProValValSerLysAlaLeuSerT 1681
27924 GFTCGCGCGCTGTGTGATCGCGCAGGCTCTCGAC..... 27960
1681 hrSerLeuValCysCysLeuCysGlnAsnProAlaAsnPhelysasp 1697
27960 ..... 27960
1698 LeuGlyAspLeuCysGlyProTyrTyrProGluHisCysLeuProIleLy 1714
27961 ...GGCGATCTCTCGCGCGCAGCTCTCGAGCGGAT..... 27995
1714 sLysProLysLeuGlyGluLysValArgProGluGlyThrCysGluGluA 1731
27996 .CGACGCGCTCTT.....CGCCCCCAAGATCGATCGCGCGCTGG 28032
1731 la..... 1731
28033 CACTGTGATCAGCTCACCAGATAGACCCCTTGGCGCTTCATCTCTT 28082
1732 .....SerLeuProLeuGluAr 1737
28083 CTCGTGCGTGGCGGCTCTCGCGACCTCAGGTCTCACTCCAACTAGCGG 28132
1737 gThrLeuGlyGlyProGluCysAlaAlaAlaAlaAlaThrAla 1750
28133 CTGGAGCGCTTCTCTCGATCGCTTGGCGACACCGCGC 28172
seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAA92302
seq_documentation_block:
ID AAA92302 standard; DNA; 31422 BP.
XX AC AAA92302;
XX DT 10-JAN-2001 (first entry)
DE S. avermitilis avermectin aglycon synthase DNA aveII SEQ ID NO:2.
DE KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
DE KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
DE KW agrochemical; ds.
XX OS Streptomyces avermitilis.
XX FH Key Location/Qualifiers
XX FT CDS 1..14646
XX FT /tag= a
XX FT /note= "avermectin aglycon synthase protein"
XX FT 14824..31422
XX FT /tag= b
XX FT /note= "avermectin aglycon synthase protein"
XX PN W0200050605-Al.

XX 31-AUG-2000.
XX 23-FEB-2000; 2000WO-JP01041.
XX 24-FEB-1999; 99JP-0046961.
XX (KITA) KITASATO INST.
XX Omura S, Ikeda H;
XX WPI: 2000-565458/52.
XX P-PSDB; AAB23751, AAB23752.
XX Avermectin aglycone synthase DNA and proteins encoded by all or part of
XX it for the production of avermectin and its derivatives for drug and
XX agrochemical use
XX Claim 2; Page 134-203; 314pp; Japanese.
XX The present sequence represents DNA which encodes avermectin aglycon
XX synthase proteins. Also described are: (1) polypeptides encoded by all
XX or part of the DNA; (2) expression vectors containing the DNA; (3) host
XX cells transformed by the vectors; (4) preparation of the polypeptides
XX by culture of the transformants; (5) preparation of avermectin aglycon
XX or its derivatives by culture of transformed avermectin-producing
XX microorganisms; and (6) oligonucleotides of 5-60 bases in length
XX containing sense or antisense sequences from the avermectin aglycon
XX synthase DNA. The enzymes are useful for the production of modified
XX forms of avermectin and of the intermediates in its biosynthesis, for
XX use as drugs, veterinary drugs and agrochemicals.
XX Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 other;

alignment_scores:
Quality: 470.50 Length: 2105
Ratio: 0.560 Gaps: 105
Percent Similarity: 39.905 Percent Identity: 21.235

alignment_block:
US-09-508-821B-6 x AAA92302/rev ..
Align seg 1/1 to reverse of: AAA92302 from: 1 to: 31422
22 GlnGluThrSerArgLeuGluAsnTyrArgGlnProSerGlnAlaGlyLe 38
6554 CGCAGGAGCTGGAGGAGAGACAGTCCAGGTACGAGTTCGAGGCTTGGC 6505
38 uSerCysasp.....ArgGlnArgL 45
6504 GCAGTCGGACATCTGTCGGCGAGAGACCGGTGAGGAGCGAGGAGTTCTT 6455
45 euLeuAlaLysAspTyrTyrAsnProGlnPro..... 55
6454 GCGCCATGCTTCCAGTGGGTGCTCTCTCCGACAGCATGGCGATC 6405
56 TyrProSerTyrGluGlyAla.....G1 64
6404 TTTCCTGCGCGCGCGGTGCGCGCCACACCTGTACACCCCGCGGTGG 6355
64 yThrProSerGlyThrAlaAlaAlaValAlaAlaAspLysTyrHisArg 81
6354 TGCCTCTCGCCAGTCGCTGAGTGGTGGAGAACTCTGCGCGTCTT 6305
81 ly.....SerLys 83
6304 CGCCAGGACCGACCGACGATGCTCGACACGCGCTCCGACACCAA 6255
84 AlaLeuProThrGlnGlnGlyLeuGlnGlyArgProAlaPheProGlyTy 100
6254 GCGCGCGCGAGCGCGGACTCACCACCCACACATCCGACCC..... 6210
```

```
100 rglvalgluaspserprotyr.prolylarginylalaglu 116
6209 .....CCACACACCGCCACACACCGCCACACCGCC 6176
117 SerLeuGlnAlaTrp.....GlyAlaProGlnPro.....Pr 127
6175 CCGCCCGCCGACAGCCAACTCCAGCGCCGACACACCGCCAGCC 6126
127 OProGlnProGlnProLeuProAlaGlyAlaAlaLysTyrAspGlu 144
6125 ACCCCGACACCGCCAGACACACCGCCAGCCAGCCAGCCAGCC 6076
144 sbleuMetLysThrAlaValProProSerArgGlnTyrAlaGln 160
6075 ACCCCCATCCGCGCCCGCCAGCGCCCTCTCCAAATCCATCCGATCC 6026
161 GlyAlaGlnValProPheArgThrHisSerLeuHisValGlnProPr 177
6025 TCCCACTCAACCCGACAGACACCGCCAGCCAGCCAGCCAGCC 5976
177 OProProGlnInProLeuAlaTyrProLysLeuGlnArgGlnLys 194
5975 ACCCCGCGCCACTCCA..... 5960
194 LAsnAspIleAlaSerProLeuProPheProGlnGlyThrHisPhePro 210
5959 .....CTCATCCGCCA...ACACGACAGCC 5936
211 GlnHisSerGlnSerPheProThrSerSerThrTyrSerSerValG1 227
5935 ACCCACTCAACCAATCCACT..... 5915
227 nglYglYglYglYglYAlaHisSerTyrLysSerCysThrAlaProTh 244
5914 .....GCACGACGG 5905
244 IagInProHisAspArgProLeuThrAlaSerSerSerLeuAlaProG 260
5904 CTCATCCACATCCAGATCCG..... 5884
260 YGlnArgValGlnAsnLeuHisAlaTyrGlnSerGlyArgLeuSerTyr 277
5884 ..... 5884
277 spGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnAlaLeu 293
5883 .....CGGACACACCGCCGCG 5866
294 GlnSerArgHisHisAlaGlnGlnThrLeuHisTyrGlnAsnLeuAlaT 310
5865 CAACGCCATCCATCTGTATCATCCAGCCGACACCGCCAGCCAGCC 5816
310 sTyrGlnHisTyrGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnProAspAla 327
5815 C.....ATGCCCGATGTTC 5802
327 IValArgThrProGlnGlnTyrTyr.....GlnThrPheSerPro 340
5801 GACTTCACGACCCCAACACACCGCCGCTCCCGCCAGACCGCTGCCC 5752
341 SerSerSerHisSerProAlaArgSerValGlyArgSerProSerTyr 356
5751 ATAAAGTGGAGACACCGCTGCGCTCATGATGATACCAACGATCGTAC 5702
357 SerSerThrProSerProLeuMetProAsn...LeuGlnAspPheProTy 372
5701 CCGTCCCGTGGCCCTCAACATCCATCCGCGCCAGGACACCGCCCA 5652
372 rSerGlnGlnProLeuSerThrGlyAlaPheProAlaGlyIleThrAsp 389
5651 CAGCCACAGCCCTGCGCAATCA.....CCGCGTCTGCGGACGAC 5611
389 Is.....SerHisPheMet 393
5610 ATTGCGCGCGCTCAACCACTTCAGAGCAGCCGCTCTGATTCACCGCAGCTCC 5561
394 ProLeuLeuAsnProSerProThrAsp.....AlaThrSe 405
5560 CCGGACACACCGCCAAAACCGATGACACGACGATCGCTGCGCAGACG 5511
405 rSerValAspThrGlnAlaGlyAsnCysLysProLeuGlnHisAspLysL 422
5510 CGCTCAACCAACAGCACACCCACACCTCGGCGCCAGCGCCACCATGCG 5461
422 eUpProGlnAsnLeuSerAspLeuSerLeuGlnSerLeuThrAlaLeu 438
5460 CCCC..... 5456
439 ThrLeuGlnValGlnAsnIleSerAsnThrValGlnGlnLeuLeuSe 455
5455 .....AC 5454
455 rLysAlaAlaValProGlnLysGlyValLysAsnLeuValSerArgT 472
5453 CCGTACGCTTGCACCGCGCGCTCGGAGACACCGCCGCTGCGCAGAA 5404
472 hPProGlnGlnHisLysSerGlnHisCysSerProGlnGlySerGlyTyr 488
5403 GTTCACAAAGCAGCCCGCGCTG...ACATCACG..... 5372
489 SerAlaGlnProAlaGlyThrProLeuSerGlnProProSerSerThrPr 505
5371 .....TGACACCGCCCGCCAGCGACGACATCCCGCCAGCTCAACGCC 5325
505 ogInSerThrHisAlaGlnPro.....GlnGlnAlaAspTyrLeuS 519
5324 TGACAGCGCAAGATGACAGCGCCACACAGACAGAAACGCGCATTCAC 5275
519 erGlySerGlnAspProLeuGlnArgSerPheLeuTyrCysAsnGlnAla 535
5274 CGTACCGCGCGACCT.....CAGGC 5252
536 ArgGlySerProAlaArgValAsnSer..... 544
5253 CAGAGGTGAGCGACGCGTCCGAGTGCAGCGCTGTGACAGCGCGCTC 5202
545.....AsnSerLysAlaLysProGlnSerValSerThrCysSerValThrs 560
5201 ATGACGCTAGCCCTGAGATCTCTGCGGCGCGGCGACAGATCGGCTA 5152
560 erProAspAspMetSerThrLysSerAspAspSerPheGlnSerLeuH 576
5431 GTCTGACTGCACATGCGCGGACAGACACCGCTGTGACCGCGCAG 5102
576 sGlySerLeuProLeuAspSerPheSerLysPheValAlaGlyGlnArg 593
5101 TGGGAGGAGTCGATGCCCGCGCTCCACAGCTCCAGGACACCTTCAGC 5052
593 spCysPro.....ArgLeuLeuLeuSerAlaLeuAlaGlnGlnAspLeu 607
5051 ATCAACGCTGCTGCGATCCATCCAGCGCTTACCGGATATCC 5002
608 AlaSerGlnIleLeuGlyLeuGlnGlnAlaIleGlyGlnLysAlaAspL 624
5001 GAAACACCGCGATGAGTACCTCGCGGACCTCCAGAAACCGCCGCG 4952
624 ySalatrpAlaGlnAlaProSerLeuValLysAspSerSerLysProPr 640
4951 GCGATATCGACGAAACCGCGCGCTCGCGATCATGAAAGCGCTCC 4902
640 oPheSerLeuGlnAsnHisSerAlaCysLeuAspSerValAlaLysSer 657
4901 A..... 4901
657 IatrpArgArgProGlnGlnProGlnAlaLeuProAspSerLeuGlnLeu 673
```

4900 ..CGTCCACACCCGGTGCACCGGAAATCTCCACCGCAT..... 4862  
574 AspLysGlyGlyAsnAlaLysAspPheSerProGlyLeuPheGluAspPr 690  
4862 ..... 4862  
690 oSerValAlaPheAlaThrProAspProLysLysThrThrLeuProLeus 707  
4861 ..... 4861  
707 erPheGlyThrLysProThrLeuGlyValProAlaProAspProThrThr 723  
4840 ..... 4840  
724 AlaAlaPheAspCysPheProAspThrThrAlaAlaSerSerAlaAspSe 740  
4804 AACGGCAGCCGCATCC...CCACAAATCGCAATCGCGTCGCAACGTCGACA 4758  
740 rAlaAsnProPheAlaThrProGluGluAsnLeuGlyAspAlaCysProA 757  
4757 CGGAGCCCG...GGACCGGAGGAGCAATGTCACGCTGGCCGCCCGC 4714  
757 rGtrpGlyLeuHisProGlyGluLeuThrLysGlyLeuGlu.GlnGlyG 773  
4713 GC.....CCTCTCCAGCTGTCTCTTGAGTATCCGGCCAGCGCG 4673  
773 yLysAlaSerAspGlyLeuSerLysGlyAspThrHisGluAlaSerAlaC 790  
4672 A.....GGAGTGGGGTATCGAATACGCT..... 4645  
790 ysLeuGlyPheGlnGluAspProProGlyGluLysValAlaSerLeu 806  
4644 .....GGTGGCAGGAGG.....CCGGGTGAGCGCTTGAGCGGTT 4606  
807 Pro.....GlyAspPheLysGlnGluValGlyCl 817  
4605 GCGCAGTTCAGCGCGCTCAGCGAGCAAGCCAGGTCCGGAAGCTC 4556  
817 yValysGluGluAlaGly.GlyLeuLeuGlnCysProGlu...ValAla 832  
4555 GCTCAGGCGTACGGGGTGGGGTGGTG...TGTCGAGCAGCGTCGCC 4509  
833 LysAlaAspArgTrpLeuGluAspSerArgHisCysCysSerThrAlaAs 849  
4508 GCGACGTACGAGCAGGTCCAGAGCGCACGTTCTCTCGCGGTGTC 4459  
849 pPheGlyAspLeuProLeuLeuProProThrSerArgLysGluAspLeuG 866  
4458 CATGCC.....TTGAGCCGCTCGGAGAACGAGT 4430  
866 luAlaGluGluGlyTrsSerLeuCysGluLeuLeuGlySerProGlu 882  
4429 CGGGGGAT..... 4422  
883 GlnArgProGlyMetGlnAspProLeuSerProLysAlaProLeuLeuCy 899  
4421 ...CGGTGGGGTTCGAGTCCGTTTCGCGGAG...GGGTG 4379  
899 sThrLysGluGluValGluValLeuAspSerLysAlaGlyTrpGlyS 916  
4378 CTTCCGGGATGTCCTGATGAGGGCGGAGTCCGGAGCCGGGAGGGA. 4330  
916 erProCysHisLeuSerGlyGluSerValIleLeuLeuGlyProThrVal 932  
4329 .....GTTGGCGGTGATCGTCCAGTC.....GATGTC 4300  
933 GlyThrGluSerLysValGlnSerTrpPheGlu..... 943  
4299 GGGACCTTCACAGGCTCTGTCATGTCTCCAGCGCTGGCCAGTGCCA 4250  
943 ..... 943  
4249 CCAGCGCGCTCCGGGTCTATCGCGCCAGACCCGCGCATCTGC 4200

944 ...SerSerLeuSerHisMetLysProGlyGluGluGlyProAspGlyG 959  
4139 CCACAGCGCCCTCGGCTATCCCTCCACACGCGGACCCACGCGCAC 4150  
959 uArgAlaProGly.....AspSert 966  
4149 GCGCACCACCGCGCGCTCACCACGCGGTCGCGAACGATTGCTCCA 4100  
966 hrThrSerAspAlaSer.....LeuAlaGlnLysProAsnLys 978  
4099 CATACGGGTTCGCGCGCTTAATCTCCCTGTCGCGCGCGCGAACGTC 4050  
979 ProAlaValProGluAlaProLleAlaLysLysGluProValProArg 995  
4049 GCGCAGCGAGGAGACACACGACACCCCGGAAAGATCCGCGCCCGC 4000  
995 yLysSerLeuArgSerArgArgValHis.....ArgGlyLeuPro. 1008  
3999 CAATCATGCAATTCAGGCGCGCGCGCTTCGCGCGAGCAGCCCGCG 3950  
1009.....GluAlaGluAspSerProCysArgAlaProValLeuProLysAsp 1023  
3949 TGACACGCTCGGAGGACAACTCCCAACACCGCG.....TCA 3912  
1024 LeuLeuLeuProGluSerCysThrGlyProProGlnGlyGlnMetGlu 1040  
3914 ATCCAACTCCCGCGCATGACACACACCGCGGCGGCTCCGCCCGG 3862  
1040 y.....AlaGlyAlaProGlyArgGlyAlaSerG 1050  
3861 AACGCGCGCGCAACACCTCCGCCAACGCTCAGGTCGCGCGCATCAC 3812  
1050 luGlyLeu..... 1052  
3811 AGCGCACACGACACCGCGCGCCCAAGCCCATCAAGTCCGCTCGGAGT 3762  
1053 .....ProArgMetCysThrArgSerLeu... 1060  
3761 TCTTCGACTCCTGGGGCTCTCCCGCGGCTGCTCAGCAGCAGGTG 3712  
1061 .....ThrAlaLeuSerGluProArgThrP 1069  
3711 TTCGCGCGCACCGCGGCGCATCCACCGGCGAGCTGCGCACCCACTCGC 3662  
1069 roGlyProProGlyLeuThrThrThrProAlaProProAspLysLeuGly 1085  
3661 CGGTGCTCTCGGTGACGAGTACGCTGCGCGG.....GGC 3627  
1086 GlyLysGlnArgAlaAla.....PheLysSerGlyLy 1096  
3626 CGCCACTCCGCTCCGCGAGCGGCTCTCCACGCGCGCGCGCACCAAC 3577  
1096 sArgValGlyLysProSerProLysAlaAlaSerSerProSerAsnProA 1113  
3576 CCGCACAAACGCCCGGAGACCGCGGCAAACTCACTCAGCCCTC 3527  
1113 laAlaLeuProValAlaSerAspSerProMetGlySerLysThrLys 1129  
3526 CCGCCACACCGCGGACACCTTACCAACCCATCGACCCCGCTCA... 3480  
1130 GluThrAspSerProSerThrProGlyLysAspGlnArgSerMetIlele 1146  
3479 TCCAGAGTCCGGCATCATCACACCGCGCGGCTCGGTCGCTC 3430  
1146 uArgSerArgThrLys.....ThrGlnGluPhe...H 1157  
3429 CGCCCCCACCACCGCGCGAGCCCCCAGCAGCAGCCGCGGCGCGCC 3380  
1157 isSerLysArgArgProSerGluGlyArgLeuProAsnCysArgAla 1173  
3379 ACAGCATCCCGCGCGCGCGGCGCGGCGCGCGCGCGCGCGCGCG 3330

2454 CATGGTCTCCGCGGGGAGGGGGCGCCGGTGAACCTTGGAGACGACGGGAA 2405  
|||||  
1373 rProVal1G1Val1G1nG1uG1Leu..... 1381  
||||| |||||  
2404 TCCGCGGGGGCGCTGAAAGGTACCTCTCCGGCCCGCGGGGAGAGTGGCCG 2355  
|||||  
1382 .....ValAsnVal1G1yThr1G1yGln1ySLeu 1390  
|||||  
2354 AACATGGGCTGCATGTGGCGGACAGTGGAAAGGCTGGCTGGTCCGACACCG 2305  
|||||  
1391 ProThrSer1yG1aLasp.....ProLeuCyArg..... 1400  
|||||  
2304 CCGGGTGGCGGGCTGTCCGTCCGGCCACCATTTGGTGGGAGTCCAGACCG 2255  
|||||  
1401 .....AsnProThrAsnArgSerL 1407  
|||||  
2254 CGTCTCGTCCCGGAGAGAGACGATGCACGCCCGGCCCTTACCGCGCGCG 2205  
|||||  
1407 eu1ySg1yLys.....LeuMet AsnSer1yS1yLeuSerTh 1420  
|||||  
2204 TGGCGACACCGGAGATGGCGATTGGTGGGACGCGGAGAGATCTCTCTC 2155  
|||||  
1420 rAsrCySPhel1yThr1G1n1A1PheThrSerProGln1aLeu1nPro 1436  
|||||  
2154 GGAAGCGCTCGATGGCCCATGGCTCCCGCGGACGGAGCGCTTCACTCA 2105  
|||||  
1437 G1yG1yThr1a1Leu1a1Pro1yS1yArgSer1yG1yValArg1aG1 1453  
|||||  
2104 GGGGGCTGTGTCCAGACCAACAGTGGCCACCGCGTGGGAGACGAGACATC 2055  
|||||  
1453 Y.....AlaHisG1yLeuSer1yS1yPro1eGln1yS1yArgPro1yT1eG 1459  
|||||  
2054 CCGGCGACATGGG.....CGGCGCGCAATTTACCGACGAGATGGCGGAG 2011  
|||||  
1459 1yPro1a1e1u1e1u1eThr1Pro1rAsp1aG1y1a1a1SerG1yThr..... 1483  
|||||  
2010 GACCTAGTGTGGCGGTGCACACCCAGGTCTTCACGACGCGGAGACGGCCCA 1961  
|||||  
1404.....GlnG1y1a1SerG1u1a1Asp1a1n.....SerG1 1492  
|||||  
1960 CCGGAGAGCGACACAGGCGGGGCTGGGCGCAACCGGTGTCTGTGCATGAC 1911  
|||||  
1492 yG1yG1yG1yS1yS1yPro1yS1yMetG1n1e1u1eG1yPro1a1a1SerG1n 1509  
|||||  
2310 CGGCTTCGGGAGAAAGTCTGGCGGTGGGAGAGATAGTCCGCGACGCCAG 1861  
|||||  
1509 rOPrOG1uG1yArg.....ProCyS1nProG1nThrArg1aG1n1yS 1522  
|||||  
1860 GGCACCGCGGGTCCGTCCGGCGGGGTGGCGGTCCGCGCATGTCG1CA 1811  
|||||  
1524 G1nProG1n1yS.....ThrAsn1ySerSer1ySer1yArg1yS1yArg1e 1535  
|||||  
1810 TGGCGTGGGAGAGAGGGGT1aCCCTCTGACAGTTCGCGGCCAGTCC 1761  
|||||  
1540 Thr1Arg1y.....Arg1a1yAsn1Thr1ThrSerSerPro..... 1551  
|||||  
1760 GCGGCGTGGGTCCCTCGCCCGGAGAAAGTACGAGGAGTGTGCGCCGAGGT 1711  
|||||  
1552 .....CyS1yS1y1yArg1a1yS1yArg1yArg1n 1566  
|||||  
1710 GGTTCGTCCCTGCACACACCGCGGAGCGGGCGCGCGCGCGAGTGGT 1661  
|||||  
1563 G1nVal1e1uPro1e1uS1yPro1a1b1yProG1n1a1e1Arg1e1u1yT1yT1 1571  
|||||  
1660 CGAGTGGCGTGGAGAACTGTGCGGTCTTGCGCCACGACACCGCAC.. 1611  
|||||  
1579 eSerSerCyS1yS1yArg1e1u1ArgSerAspSerArgThrPro1a1PheSer 1594  
|||||  
1612 .....GAGTGTCCGACACGACGAC...GCTC 1591  
|||||  
1596 rOPheVal1ArgVal1G1u1yS1yArgAsp1a1a1PheThr1Thr1Le1yS1yThrVal 1611  
|||||



XX 02-MAY-1997; 97EP-0107329.  
 PF  
 XX  
 PR 07-MAY-1996; 96US-0016753.  
 XX  
 PA (HOFF) HOFFMANN LA ROCHE & CO AG F.  
 XX  
 PI Reeves CD, Soliday CL;  
 XX  
 DR MPI: 1997-538619/50.  
 DR P-PSDB: AAM34139-W34219.  
 XX  
 PT Streptomyces frenolicin gene cluster - useful for producing  
 PT recombinant frenolicin antibiotics  
 XX  
 PS Claim 1: Page 40-60; 66pp; English.

This DNA sequence comprises the Streptomyces frenolicin gene cluster containing specifically claimed coding sequences (genes A-U) that respectively encode 21 proteins (see AAM34199-219) involved in frenolicin synthesis. The genes can be divided into 5 subclusters: (1) genes A, B, C, D and E encode an efflux pump; (2) genes H, I, J and K encode butyrate starter synthases; (3) genes L, M and N encode polyketide synthases (PKS); (4) genes O, P, Q and R encode a hemiketalase, a ketoreductase and cytochromes/oxidases; (5) genes S and T encode a keto/enoyl reductase and a hydrolase. Also claimed are vectors, host cells (especially a Streptomyces sp., particularly Streptomyces roseotulvus), and the encoded proteins. Cells transformed using the above sequence can be cultured to produce frenolicin or frenolicin precursors. The methods. The frenolicin can be oxidised to frenolicin B, an antibiotic used as an anticoccidial agent. The frenolicin can be used as animal feed additives.

Sequence 24379 BP; 3077 A; 9792 C; 8499 G; 3011 T; 0 other;

alignment\_scores:  
 Quality: 468.50 Length: 2125  
 Ratio: 0.571 Gaps: 104  
 Percent Similarity: 38.588 Percent Identity: 21.224

alignment\_block:  
 US-09-508-821B-6 x AAT93095 ..

Align seg 1/1 to: AAT93095 from: 1 to: 24379

22 GINGLURHRSERARGLEUHSNTYRARGLN.....Prose 34  
 |||||.....  
 11835 CAAAGAGCTGGTCGACGAGGCTACGACGCGGAAATCGCGCAAGCCCTCC 11884  
 34 rGINLAAGLYLEUSERCYSPARGLNARG.....LeuLeuAlaL 48  
 :|||:.....  
 11885 ACGGCGAGGCCCTCGACACGAGCTTCAACCAATCGGGTTACACATCGCTG 11934  
 48 YSASPRTYRASNPROGLNPROTYRPROSER..... 58  
 :|||:.....  
 11935 ACGGTCTACGAGATCGTACACCGCATCCAGAGACGAGACGAGTACCGGT 11984  
 59 TYRGLUGLYGIALAGLYNTPROSERGLYTHRALAALAAVALAAL 75  
 :|||:.....  
 11985 GCCCGACGAGAGCTGAGACCTCTCGACACCCCGCGCCCTTCATCGGCT 12034  
 75 AASPYSITYRHISARGLYSELYSALALEUPROTHGLINGLYLEUG 92  
 :|||:.....  
 12035 ACGTCGACGCGCGCGGTTCCCGACCTACCCGACGAGGAGGT...G 12081  
 92 INGLIARG.....ProAlaPheProGlyTYRGLYVALGlnAspSerSer 106  
 :|||:.....  
 12082 CGGGAGAGGCGGGGCGGAGCGGGGCGGCGGCGGTACGAGCGCG 12131  
 107 PROTYRPROGLYARGTYRALA..... 113

12132 CCGGCGCCCGGCGGACGCGCTCACGCCCGCTCCGTACGACCTGACCC 12181  
 |||||.....  
 12134 .....GlyLeuG 116  
 12182 GCGTGTGAGAGCGCGCGCAAGAGCCCGACAGCTGTCGCGGGGTCCGC 12231  
 :|||:.....  
 12232 GGGACGCGCGCGGAGAGACCGACAGACGAGGCGCGCGCCGACGAGACG 12281  
 133 PROLEUPROBLAAGLYVALAALYSYRASPGLUAENLEUHEULYSLYSTH 149  
 :|||:.....  
 12282 CCGAGATACGCGCGCGCGGCGCGGACGACCTCCGCGAGAGGCTCG 12331  
 149 VALAVALPRO.....ProSerArgLnt 157  
 :|||:.....  
 12332 AGCAGACACACAGCCACGCCCGCGCGCTGTCGCGCGCTCCGCTCCG 12381  
 157 YRALAGLUGLNGLYALAGLVALPROHE.....Arg 167  
 :|||:.....  
 12382 CGTCGCGAGAGAGTTCGCCAGCGCGCGGCGACGACGAGGCGGCGTCA 12431  
 168 THNHSERLEUHSVALGLNGLNPROPROPROPROGLINGLNPROLEUAL 184  
 :|||:.....  
 12432 CGTGACCGCACACAGCGCGCGCGCGACCGCTCCGCGCGCGGTAGCGG 12481  
 184 ATYRPROYSLEUHLNARGLYNLYLEUGLNASNAPLLEALSERPROL 201  
 :|||:.....  
 12482 CGACCGCGTCCCGCGCGACGACAGAGGCGCGGCGGCGCGACCGCGC 12531  
 201 EUPROHEPROGLINGLYTHNHSPIHEPROGLNHISERLINSEPHPRO 217  
 :|||:.....  
 12532 GACCGCTTCCGCGCTCG.....CCCTGTCACAGCGCGCGCGCGC 12572  
 218 THNSERSETHYRSESRSERYALGNGLYGLYGLNGLYALAH 234  
 :|||:.....  
 12573 CGCGGAGGCGGGGAGTGAAGGTACCTCCGCTCGAGGCGGCGCGC 12622  
 234 SERYTYRLYS.....SerCyTHRALAPROT 243  
 :|||:.....  
 12623 GTACGCCCTGCTCGCGCAAGCGCGCGCGCACCGCGCGCGCGCTC 12672  
 243 HVALAGLN.....ProHisASPArgProLeu...ThrAlaSerSerSe 256  
 :|||:.....  
 12673 CGGCGCGCACACGACGCTGACGCGCGCGCTACCGCGCGCCACACCA 12722  
 256 rLEUALAPROGLYGLNARGYALGlnASNLEUHSALTYRGLNserGLYA 273  
 :|||:.....  
 12723 CGTGGCGCCCGACGATACGCGGCGAGTCTCCACCGCGCGCGCGCGAG 12772  
 273 rGLAUSERTYRASPGLNGLNGLNGLNGLNGLNGLNGLNGLNGLN 289  
 :|||:.....  
 12773 AGCATGCGCCCGGAGCGGCGCGCGGAGACCGGCTCGTCCGCGAG 12822  
 290 GINGLINALLEUHLNserArgHISHSALAGLNGluthrLeuHISLYRPL 306  
 :|||:.....  
 12822 CAGCGGAGCGCGCGCGCGCGCGC..... 12846  
 306 rASNLEUALLYSTYRGLNHISLYRGLYGLNGLNGLNGLNGLNGLNGLN 323  
 :|||:.....  
 12847 .....GCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 12873  
 323 LNPProASPALAVALARGTHNPROGLNGLNGLNGLNGLNGLNGLN 338  
 :|||:.....  
 12874 .....GAGTTCGCGAGCGTGTGCGCGAGCAAGAGTGGCGGACCGCTG 12918  
 339 SerProSerSerHisSerPro..... 346  
 :|||:.....  
 12919 CCGCCCGCGCGCGAGCGCGCGCGCGAGCGCGTACCCGACGAGGAGCA 12968  
 347 .....AlaArgSerValGlyArgSerProSerTYRser 357  
 :|||:.....

12969 ACGGCTGCGGACAGCGGCGGCTTCCAGGCGCTCTCTCGCGCGCGCGCG 13018  
358 SerThr.....ProSerProLeuMetPro..... 365  
13019 TCAGCCACCGCGCGCGGCACTCTCCCGCCAGGTGCCACACAGCGGAA 13068  
366 .....AsnLeuGluAsnPheProThrSerGlnG 375  
13069 GCATCTCTCCACCGCGGCTGTTGAACTCCGGCCACCGCGGTGAGTCCGG 13118  
375 InProLeu.....Ser 378  
13119 CGGCATCCGCGCGCGCTCGCGCGCTGACCGCGGAGCAGCGGCGCAC 13168  
379 ThrGlyAlaPheProAlaGlyLeuThrAspHisSerHisPheMetPro.L 395  
13169 ACGGGGCGCGCTCTCTCTCGTACCGCTCGCTCCGACACTTTCGCGC 13218  
395 euLeuAsnProSerProThrAspAlaThrSerSerValAspThrGlnAla 411  
13219 TCCTGCTCCG...CGACTCTCTGCC.....GGTACCGAGCCT 13253  
412 GlyAsnCysLys.....ProLeuGlnLysAspLysLe 422  
13254 CGCCTTCGCGCTCGACGCGCGTGGACTCCCGCTCCAGCGCGC.... 13298  
422 uProGluAsnLeuSerAspLeuSerLeuGlnSerLeuThrAlaLeuT 439  
13299 .CGCGCGCGCGCGCCACCGAGCTCAAGCCGCTTCGAGGAGCGCTTCC 13347  
439 hrLeuGlnValGluAsnIleSerAsnThrValGlnGlnLeuLeuSer 455  
13348 AGGTCGGTTCGAGCGCGTGAACGCGACC..... 13376  
456 LysAlaAlaValProGlnLysGlyValLysAsnLeuValSerArgTh 472  
13377 ...GCGCGCTCCAGGACGCGCAAGGGGTCTGTCAGTGAACCGACAATC 13423  
472 rProGluGlnHisLysSerGlnHisCysSerPro..... 483  
13424 GCGG.....TCACCGGCATCGGCGGTGTGGC 13449  
484 ....GlutylserGlyTyrSerAlaGluProAlaGlyThrProLeuSer 498  
13450 TCGGGAGGATCGGCGCAAGCGTACTGGAGCAGCTACCTCCGAGC 13499  
499 GluProProSerThrProGlnSerThrHisAlaGluProGlnGluAl 515  
13500 GCACCGCCACCGCGCGCTCTCTCTTCGAGCGCTCGCCT..... 13541  
515 aaSerTyrLeuSerGlySerGluAspProLeuGluArgSerPheLeuTyrC 532  
13541 ..... 13541  
532 ysAsnGlnAlaArgGlySerProAlaArgValAsnSerAsnSerLys... 547  
13542 ....TCGGTTCGGATCGCGCGAGTTCGACTTCGACCGCGCGCG 13586  
548 .....AlaLysProGluSerValSer..... 554  
13587 CGGGCTGAGCCCGCGAGTCCGCGCATGGACCGCGCGCGCGAGTTC 13636  
555 .....ThrCysSerValThrSerP 561  
13637 GCGTGTGAGCGCGAGGAGAGCTCGCGAGCAGCGGACTCGACGTGC 13686  
561 roAspAspMetSerThrLysSer.AsAspSerPheGln.SerLeuHisG 577  
13687 CGACTTCGACCCCGGATCGGGGTGAGCATCGCGAGCGCGCGTGGCG 13736  
577 lysLeuProAspSerPheSerLysPheValAla..... 589  
13737 GGACCACTCGCTGAA.....CGGAGTACCTCGCCCTCAGCGACAGC 13780

590 .....GlyGluArgAspCysProArgLeu.....LeuLeuSe 600  
13781 SGGGACAGTGGGAACCTGCTCTACCTTCGCGGCACTCTACGA 13830  
600 rAlaLeuAlaGlnGluAspLeuAlaSerLulLeuGlyLeuGlnGluA 617  
13831 GCGCTTACCCGCACTGCTTCGCGCGAGGTGGCGGGTG..... 13873  
617 lalLeuGlyLysAlaAspLysAlaTrpAlaGluAlaProSerLeuVal 633  
13874 ..ATCGGC.....GCGAGGCGCGCGCGCAGTC 13900  
634 LysAspSer.SerLysProProPheSerLeuGluAsnHisSerAlaCysL 650  
13901 GTCTCCACCGGTGCACCTTCGGCATCG..... 13928  
650 euAspSerValAlaLysSerAlaTrpProArgProGlyGluProGluAla 666  
13928 .....ACTGCTTCGCGCACGCGCGGAGCTCATCGCGAGGCA 13967  
667 LeuProAspSerLeuGlnLeuAspLysGlyGlyAsnAlaLysAspPheSe 683  
13968 GCGCGGAGTCTGTC..... 13982  
683 rProGlyLeuPheGluAspProSerValAlaPheAlaThrProAspProL 700  
13983 .....TCGCGGCGGCGCACCGACACCCCATCTC 14010  
700 ysLysThrThrGlyProLeuSerPheGlyThrLysProThrLeuGlyVal 716  
14011 GCGGATCGCGCTCGCTTCGACGCCATCAAGGCCA..... 14048  
717 ProAlaProAspProThrThrAlaAlaPheAspCysPheProAspThrTh 733  
14049 COTCGCCCAACAACGACGACCGCGCGCACGCTCGCGCTTCGACAGG 14098  
733 rAlaAlaSerSerAlaAspSer..... 740  
14099 GAACGCAACGGCTTCTCTCGCGGAGGCGCGCTCTCTCTGCTGGA 14148  
740 ..... 740  
14149 SGAGCTGGGCGACGCGCGCGCGCGCGCGCTCTACGCTCTCTGT 14198  
741 ..AlaAsnProPheAla..... 745  
14199 CGGCTACGCGACGCGCTGCAACGCGTACCATGACCGGCTCACCCG 14248  
746 .....TtpProGluGluAsnLeuGlyAspAlaCysProArgT 758  
14249 CACGGCAGGAGATGGCGAGGCGATCC...GGCACGCGTCCGCGAGAG 14295  
758 rpgLysLeuHisProGlyGluLeuThrLysGlyLeuGlnGlnGlyLys 774  
14296 CGGACCGACCGCGCGCTGGACTACGTCAACGCCAC..... 14335  
775 AlaSerAspLysLeuSerLysGlyAspThrHisGluAlaSerAlaCysLe 791  
14336 ...GECTCCCGCACCAACGACGACCGCGCACGACCGCGCGCTTC... 14380  
791 uGlyPheGlnGluAspProProGlyGluLysValAlaSerLeuProG 808  
14381 .....AAGCGCAGGTTCGGCGAGCGCGCGAGCGGTGGCG. 14416  
808 lyAspPheLysGlnGluGluValGlyGlyValLysGluGluAlaGly... 823  
14417 .....GTCAGCTCCATCAAGTATGATCGGCGAC 14446  
824 .....GlyLeuLeuGlnCysProGluValAlaLysAlaAspArgTirple 838  
14447 TCGCTCGTCCATCGGTGCTGGAGATCGG.....GCTGG... 14485

838 ugiuaspserarghiscyscysserthrAlaaspheglyaspLeuProL 855  
14486 .....GCGCTGCGACAGAGTACGCGC.....G 14507  
855 euLeuProThrSerArglyGluaspLeuAlaGluGluTyr 871  
14508 TGGTGCGCGCCGACGCGACCTGACAGCGCCGCGCGAG..... 14548  
872 SerSerLeuGlyGluLeuGlySerProGluGluArgProGlyMetG1 888  
14549 .....TGCAGCTCGACCTACGTGCGCCACGAG..... 14575  
888 nasProLeuSerProLysAlaProLeuIleCysThrLysGluGluValG 905  
14576 .....GCGCGCGAGAGCGCGGTG 14594  
905 IuGluValLeuaspSerLysAlaGlyTPGlySerProCysHisLeuSer 921  
14595 ACAGGGTCTGTGCGTGGCGAGCGGCTTGCGGT..... 14629  
922 GlyLysSerValIleLeuLeuGlyProThrValGlyThrGluSerLysVa 938  
14629 ..... 14629  
938 LgInSerTrpGluSerSerLeuSerHisMetLysProGlyGluGlu 954  
14630 .....TTCAGAGCGCGCATGTTC.....TCACCGGAGACACGG 14663  
955 GlyProaspGlyGluArgAlaProGlyaspSerThrThrSeraspLase 971  
14664 GCGCCCGCTCCGAGCG.....CCTGACACCG 14692  
971 rLeuAlaGlnLysProAlaLysProAlaValProGluAlaProIleAla 988  
14693 CTCACCGCAGCGCGCGCGAGCGCGCGCG..... 14726  
988 ySLysGluProValProArgGlyLysSerLeuArgSerArgVal... 1003  
14727 .....GCCAGTCCGCGACCGCTGCGACAGGAGGACGCC 14762  
1004 .....HisArgLysLeuProGluAlaGluaspSerProCysArgAl 1017  
14763 GCCCATGACACGCGCCCTCCCGACCGCGCGAGCGCGCGCGCG 14812  
1017 a.ProValLeuProLysaspLeuLeuProGluSerCysThrGlyPr 1033  
14813 CCGCCCTGCGCGCGCTCA.....CCGACATCGCGCTGCGCGCG 14853  
1033 oProGlnGlyMetGluGlyAlaGlyAla.....ProGlyArgGlyVa 1048  
14854 CCCACGCGGCTCGCGACAGAGAGTGTGGGCGCGCGCGCGGGA 14903  
1048 Iaserglu...GlyLeuProArgMetCysThrArgSerLeuThr.... 1061  
14904 GCACGCTGCGCGCGCTCACGAGTACAGCAGCAGCGCGCGCGGTG 14953  
1062 .....AlaLeuSerGluProArgThrPro..... 1069  
14954 GACTGCTGCGCGCGCTCGCGAGTTCGACGCGCGCGCGCGCGCG 15003  
1070 .....GlyProGlyLeuThrThrPro. 1078  
15004 CGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15053  
1079 .....AlaProProaspLysLeuGlyLysGlnArg 1089  
15054 CGACGAGCACTGAAGAGCGCGCGCTGACCGCGCGCGCGCGCG 15094  
1090 AlaIaIaIaSerSerGlyLysArgValGlyLysProSerProLysAlaI 1106  
15095 ...TCCCGAGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15141  
1106 aserSerPro.....SeraspProAlaIaLeuP 1116

15142 TTCAGATCACCCACCGCGAGATCCGCAAACTCTGAGACGAGCGCGCG 15191  
1146 roValIaIaSeraspSerSerPrometGlySerLysThrLysGluThrasp 1132  
15192 CCGGTACGCGCTTCAGAGTCTTCGCGCGGTTCAG..... 15229  
1133 SerProSerThrProGlyLysaspGlnArgSerMetIleLeuArgSerA 1149  
15230 ...CCGCAACACCGCGCGAGATCCATCCGCGACGCGGAGCGCGCGCG 15276  
1149 rGlnLysThrGlnGluIlePheHisSerLysArgArgProSerGlu 1165  
15277 GCGCGCTGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 15326  
1166 Gly..... 1166  
15327 GCGCGCGCGGTACGCGAAGGCGGCGGTGCGCGGTGAGCGCGCGCG 15376  
1166 ..... 1166  
15377 TGAAGTCGCGCTGACCCCTGAGGCGTGGCGCGCGCGCGCGCGCG 15426  
1166 ..... 1166  
15427 ACCTGACGCGCTCGCGCGACCGCGCGCGCTACTCTCCCTGACCG 15476  
1166 ..... 1166  
15477 GCGTGCCTCGCGACGCTGTGCGAGAGCGCGCGCTCTCTCAACCTGG 15526  
1167 .....ArgLeuProaspCysArgAlaI 1174  
15527 AGACACCGCGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGA 15576  
1174 hrLysLysLeuaspSerThrLeuProAlaThrPheLysValSer 1190  
15577 CTCGCGGAGTA.....CGCGCGACGTTCAGACCGCGCGCGCGCG 15620  
1191 SerSerProGlnLysGluGlyArgValSerGlnArgAlaValProLy 1207  
15621 ACG 15670  
1207 sProGlyAlaGlySer..... 1212  
15671 GCGTGGACCGCGCGGACGTGACGTCTGTCGCGCGAGCGCGCGCGCTC 15720  
1213 .....LysLeuSeraspArgProLeuHisAlaLeu..... 1222  
15721 CCG 15770  
1223 LysArgLysSerAlaPheMetAlaProValProThrLysLysArgAsnLe 1239  
15771 CCGCGCGGTTCGCGTACGCGCGCGCGCGCGCGCGCGCGCGCG 15809  
1239 uValLeuArgHisGlySerSerSerSerSerAlaSerAlaMetGlyG 1256  
15810 .....GCGCGTGCCTGCGGCG 15825  
1256 IuMetGlyArgArgArgGlyLeuArgValProProSerSerArgGly 1272  
15826 GCGCGCGCGCTCGACGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 15872  
1273 CysLeuLeuProArgLysProSer...ProProArgAlaMetAlaSer. 1287  
15873 .....CCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15913  
1288 .....LeuProGlnSerSerHisProArgArgProPheLeu. 1300  
15914 CGAGCGCGCTGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15963  
1301 .....ProGlnAlaArgLeuSerSerAlaIaIaPheGlnAlaMet 1313



```
15964 ACGGCGCTCTCTCGCGCGCGCCACGGCGCTTCAACGCGCGCTCT 16013
1314 LysThrLysValLeuPro.ProArgLysGlyArgGlyLeuLysLeuGluA 1330
16014 CGTCCGCGCGCGCGCGCGCGCCCTACCGCTCAGG... 16050
1330 LaileValGlnLysIleThrSerProSerLeuLysLysPheAlaCysLys 1346
16051 .....CCT 16053
1347 AlaProGlyAlaSerProGlyAsnProLeuSerProSerLeuSerAspLys 1363
16054 GCCCGGCTGCGCGCGCG...CGCGCGCACCCCGCGCTCCACC... 16092
1363 saspArgGlyLeuLysGlyAlaGlyLysSerProValGlyValGluGlu 1380
16093 .....GCCCGCACCCCGCGCGCTCCCATGA 16123
1380 LysLeuValAsnValGlyThrGlyLysLeu.....ProThrSerGly 1394
16124 GGCACACTGCGCTGCGACCTCAAGAACTGCTCCGCGAGACCGCGGG 16173
1395 AlaAspProLeuCysArgAsnPro.....ThrAsnArgSerLeuLys 1408
16174 GAGGACGACAG...CGTGCACCTGCGCGGAGAACTCGACAGCCCTCG 16219
1408 sGlyLysLeuMetAsnSerLysLeuSerSerThrAspCysPheLysT 1425
16220 TGAACCTCGGTACGACTCCCTCGCGCTGCTGGA..... 16253
1425 hrGluAlaPheThrSerProGluAlaLeu.....Gln 1435
16254 ..GACGGCGCGCTGCTCCAGAGCGGTACGGCATCGCGTACGTGAGA 16301
1436 ProGlyGlyThrAlaLeuAlaProLysLysArgSerArgLysGlyArg.. 1451
16302 GACGGTCCGCGGTGGCGGCCCGCGCGGAACTGCTCGAGAGGTCAACA 16351
1452 AlaGlyAlaHisGlyLysSerLysGlyProLeuGluLysArgProTyrL 1468
16352 CCACCGCGCCACCGCTGAGC..... 16373
1468 euGlyProAlaLeuLeuLeuThrProArgAspArgAlaSerGlyThrGln 1484
16374 .....GGCGGCGCACGCGAAGAG 16394
1485 GlyAlaSerGluAspAsnSerGlyGlyLysLysProLysMet.. 1500
16395 GGTCCGCGCCACCCCGCGCGGCGGTGGGCGGACCTTCGTTCGC 16444
1501 .....GluLysLeuGlyProAlaSer..... 1507
16445 GCGGACCTCACAGAGGTGCGTCTGGGTACCTGCGCGACAGCT 16494
1508 .....GlnProGluGlyArgProCysGln..... 1516
16495 GCGAGTCCCGCGCTCCCGACCGAGCTCCCGTACGACGAACTGG 16544
1516 ..... 1516
16545 GACGATCACGGGAGCGCGCTCCGCGCGGTCAACACGACGCGG 16594
1517 .....ProGlnThrArgAlaGlnLysGlnProGlyHisThrAsnTyrS 1531
16595 TAGAACTCCGCTCGACGCGCGCTCCACCGGCGGTACGTCCAGCAT 16644
1531 exSerTyrSer.....LysArgLysArgLeuThrArgGly... 1542
16645 CGAGAACCAAGTCCGCGCTGAGCGGGTGTCTGGGGAACCGGGGCGGT 16694
1543 .....ArgAlaLysAsnThrThrSerSerPro..... 1551
16695 TGAAGCGCGCACCTCTCCGCGATCGCGCGCTCCCGCGCGCGCGG 16744
```

```
1552 .CysLysGlyArgAlaLysArgArg.....ArgG 1561
16745 GTGTGGGGGAGTGGCGAAGAGCCCTCCCGCCGCAAGGTGGCGCGAA 16794
1561 LncLncLncValLeuProLeuAspProAlaGluProGluLysLeuLys 1577
16795 CTCTCGCGCGGTACCGCTCCAGGTGGTGGCTGGCGCGGTAGAAG 16844
1578 TyrIleSerSerCysLysArgLeuArgSerAspSerArgThrProAlaP 1594
16845 CCGTACTCTCGGTAGAGCGCGCGGAACGCGCGCTCGCGCGCGAGG 16894
1594 eserProPheValArgValGluLysArgAspAlaPhe..... 1606
16895 GACC.....CGGCCCTCACGTGCTCGCGCGGTGGTGGTGGTGCAC 16935
1607 .....ThrThrIleCysThrValValAsnSerProGly..... 1617
16936 GGAGAGCTCCAGAGGTGT.....CTGACCTGTCCGCGCAAGCTCCG 16979
1617 ..... 1617
16980 CGCGTGGTCTGACGGCGCTTGAGCCCTCAAGGAGCGCGCGGAGG 17029
1618 .....AspAlaProLysProHisArgL 1625
17030 CTGGCACACGACCGCGCGTACACCGACCCCGCGCGCGGAGGA 17079
1625 ysProSerSerSerAlaSerSerSerSerSerSerSerSerPheSerLeu 1641
17080 AGGCCATCATGACACA..... 17096
1642 AspAlaAlaGlyAlaSerLeuAlaThrLeuProGlyGlySerIleLeuGl 1658
17097 .....GCAGCTCCGACACCGCGCGCGGAGGAGCGACAC 17134
1658 nProArgProSerLeuProLeuSerSerThrMethIleLysGlyProValV 1675
17135 CGGGGACCGCTCTCGTGACCGCGCGCGCGGCGGTATCGGCTCGCGG 17184
1675 alSerLysAlaLeuSerThrSerCysLeuValCysCysLeuCysGlnAsn 1691
17185 TCGCGGAGGCGCTCTGCGCGCGGCTATCCCGTGGTGGTCTGCGCGCG 17234
1692 ProAla 1693
17235 GACGCG 17240
seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:AAV25925
seq_documentation_block:
ID AAV25925 standard; cDNA; 24379 BP.
XX
AC AAV25925;
XX
XX 15-JUL-1998 (first entry)
XX
DE Streptomyces roseofulvus frenolicin gene cluster.
XX
KW Streptomyces roseofulvus; frenolicin gene cluster; frenolicin B;
XX antibiotic; ss.
XX Streptomyces roseofulvus.
XX
XX Key Location/Qualifiers
XX CDS 636..2948
XX /tag= a
XX /trans_except= (pos:636..638,aa:Met)
XX /note= "encodes protein given in AAV55800"
XX CDS 2945..3916
XX /tag= b
XX /trans_except= (pos:2945..2947,aa:Met)
XX
```





617 LaIleGIyLysAlaAspLysAlaTrpAlaGluAlaProSerLeuVal 633  
 13874 ..ATGGCC.....GCGAGAGGGCCCGGCAATC 13900  
 634 LysAspSer.SerLysProProPheSerLeuGluAsnHisSerAlaCysL 650  
 13901 GTCTCCACCGGGCTGCACCTCCGCAATC..... 13928  
 650 euAspSerValAlaLysSerAlaTrpProArgProGlyLeuProGluAla 666  
 13929 .....ACTGGCTGGCCACGCCCGGCACTCAACCGCGAGGCA 13967  
 667 LeuProAspSerLeuGluLeuAspLysGlyValAsnAlaLysAspPhe 683  
 13968 GCCCGCAACCTGTCG..... 13982  
 683 rProGlyLeuPheGluAspProSerValAlaPheAlaThrProAspProL 700  
 13983 .....TCGGGGGCGCACCGACACCCCGCATCTC 14010  
 700 yslYsrThrGlyProLeuSerPheGlyThrLysProThrLeuGlyVal 716  
 14011 GCCGATCGCGCTGCGCTCTTCGACGCCATCAAGGCCA..... 14048  
 717 ProAlaProAspProThrThrAlaAlaPheAspCysPheProAspThrTh 733  
 14049 CCTGCCCCACAGACAGACGCCCGCACGCTCCGCGCCCTTCGACAG 14098  
 733 rAlaAlaSerSerAlaAspSer..... 740  
 14099 GAGCGACAGGCTTCCTCTCGAGAGGCGCGCCGCTCCCTGCTGGA 14148  
 740 ..... 740  
 14149 GAGCTGGGGGACGCGCGCGCGCGACGCCACCTCACTGCTCTCTGT 14198  
 741 ...AlaAsnProPheAla..... 745  
 14199 CCGGCTACGCCACGCGCTGCACAGCGATACACATGACCGGGCTACCCCG 14248  
 746 .....TrpProGlyLeuAsnLeuGlyAspAlaCysProArg 758  
 14249 CACGGCAGAGAGATGAGCGAGGCAATCC..GGCAGCGCTCGCGAGAG 14295  
 758 rPGLyLeuHisrProGlyGlyLeuThrLysGlyLeuGluGlyGlyLys 774  
 14296 CAGGACCGACCCCGCGCGTGAATGATCAAGGCCAC..... 14335  
 775 AlaSerAspGlyTlleSerLysGlyAspThrHisGluAlaSerAlaCysLe 791  
 14336 ...GGCTCGGGCACCAAGCAGACGCGACGAGACGCGCGCTTC.. 14380  
 791 uGIyPheGluGluGluAspProProGlyGlyLysValAlaSerLeuProG 808  
 14381 .....AAGGCCACGCTGCGGCGACGGGCGAGAGGCGCG.. 14416  
 808 LysAspPheLysGluGluValGlyValGlyValLysGluGluAlaGly.. 823  
 14417 .....GTCACTCGATCAAGTCAATGATCGGCAC 14446  
 824 .....GlyLeuGluGlyCysProGlyValAlaLysAlaAspThrTriPle 838  
 14447 TCCTCGTGGTCATCGGTCGTGGAGATCGG.....GCTCGG.. 14485  
 838 uGIAspSerArgHisCysCysSerThrAlaAspPheGlyAspLeuProL 855  
 14486 .....GCGCTCGCATGGAGATAGGC.....G 14507  
 855 euLeuProProThrSerArgLysGluAspLeuGluAlaGluGluGlyTr 871  
 14508 TGTGTCCGCGCCACGCGCAACCTGACACGCCCGGACCCCGAG.. 14548

872 SerSerLeuGlyGluLeuGluGlySerProGluGluArgProGlyMetG 888  
 14549 .....TCGACCTTGACTACGTGCCCAAG.. 14575  
 888 nasProLeuSerProLysAlaProLeuIleCysThrLysGluGluValG 905  
 14576 .....GCGCGGAGACCGGGGTGG 14594  
 905 LysGluValLeuAspSerLysAlaGlyTrpGlySerProCysHisLeuSer 921  
 14595 ACAGGTCCTGAGCGCGACGCGCTGGCGCGGT..... 14629  
 922 GlyLeuSerValIleLeuGluGlyProThrValGlyThrGluSerLys 938  
 14629 ..... 14629  
 938 LysSerTrpPheGluSerSerLeu.SerHisMetLysProGlyGluGlu 954  
 14630 .....TTCCAGAGCGCCATAGTCC.....TCACCGGAGACACG 14663  
 955 GlyProAspGlyGluArgAlaProGlyAspSerThrThrSerAspAlaSe 971  
 14664 GCGCCGCGCTCCGACG.....CTGACACCG 14692  
 971 rLeuAlaGluLysProAsnLysProAlaValProGluAlaProIleAla 988  
 14693 CTACCGGACGCGCCCGACCGCGCGGCGCGG..... 14726  
 988 yslYslGluProValProArgGlyLysSerLeuArgSerArgVal... 1003  
 14727 .....GCGAGGTCCCGCACCGCTGCACAGGAGGACGCC 14762  
 1004 .....HisArgGlyLeuProGluAlaGluAspSerProCysArgAl 1017  
 14763 GCCCATGACCAACCGCCCGCCGACCGCCAGCGCCCTCCCGGGG 14812  
 1017 a..ProValLeuProLysAspLeuLeuLeuProGluSerCysThrGlyPr 1033  
 14813 CCGCGCTGCCCGCGCTCTCA.....CCGGCATCGCGCTCGCGCC 14853  
 1033 oProGluGlyGluMetGluGlyAlaGlyAla.....ProGlyArgGly 1048  
 14854 CCGACAGGGCTCGGACGAGAGAGTGTGGGCGCCACCTGCGGGGGA 14903  
 1048 LasSerGlu..GlyLeuProArgMetCysThrArgSerLeuThr.... 1061  
 14904 GCACGGTGTGGCGCGCTGACGAGTACAGCGAGCGGCGACCCCGGTG 14953  
 1062 .....AlaLeuSerGluProArgThrPro..... 1069  
 14954 GACTGTGCGCGCGCTCCGGAATTCAGACCGCGCGCACCGCCGCG 15003  
 1070 .....GlyProProGlyLeuThrThrThrPro.. 1078  
 15004 CGCTCTGCGCGACAGACGACCGGGTACCCGGCTCGCGCTCGCGCC 15053  
 1079 .....AlaProProAspLysLeuGlyLysGluArg 1089  
 15054 CGACGAGCACTGAGAGAGCGCGCTGACCGCGCGG..... 15094  
 1090 AlaAlaPheLysSerGlyLysArgValGlyLysProSerProLysAlaAl 1106  
 15095 ...TCGCCGAGTACGGGGGAGACCGCGCTCACTCCACAGCCACCGGGG 15141  
 1106 aserSerPro.....SerAsnProAlaAlaLeuP 1116  
 15142 TTCAGATCACACACGACGAGATCCGAACTGTGACGAGGCGCGG 15191  
 1116 roValAlaSerAspSerSerProMetGlySerLysThrLysGluThrAsp 1132  
 15192 CCGCGTCAAGCTTACAGAGCTTCGCTGCTCTTCG..... 15229  
 1133 SerProSerThrProGlyLysAspGluArgSerMetIle..LeuArgSer 1149

15230 .....CCTCAACACGGCGCAGATCTCCATCGGACACGGCATCGCGGCC 15276  
1149 rgthrLysThrGlnGluIlePheHisSerLysArgArgProSerGlu 1165  
15277 GCGCGCGTCGTCGTGGCCGACGACGCGGCGGCTCGACGCTCGGCCA 15326  
1166 Gly..... 1166  
15327 GGCCGCGCGCTACTGCGCAAGGCGGGGTCTGGCGGTGAGCGGCGGG 15376  
1166 ..... 1166  
15377 TGGAGTCCGCGCTCGACCCCTCGGGCCTGGGCCCGCCACGCGCTCTCGGGC 15426  
1166 ..... 1166  
15427 ACCTAGCGCGCTCCGGCGACCCGCGCCACCGCTACCTCCCTTCGACCG 15476  
1166 ..... 1166  
15477 GCGTGCCTCGGCACGCTGTCGCGAGGCGCGCCCTCTCTCACCTCG 15526  
1167 ..... ArgLeuProAsnCysArgAlaT 1174  
15527 AGACACCCGCGCACCGGAGCGCGACGCGACCCCGGATCTACGGCGAA 15576  
1174 hrLysLysLeuLeuAspAsnSerHisLeuProAlaThrPheLysValSer 1190  
15577 CTGCGCGGTA.....CGCGCGCAGTTCGACCGCGCGGCTCCGG 15620  
1191 SerSerProGlnLysGluLysArgValSerGlnArgAlaAaGValProLy 1207  
15621 ACGGCGCGCGCTCGAAGCGCGCGCGCTCGCGCTCGCGACGCG 15670  
1207 sProGlyValLysSer..... 1212  
15671 GCTGCGACCGGGAGCTGCGCTGCTCGCGACCGCGCGGGCTC 15720  
1213 .....LysLeuSerAspArgProLeuHisAlaLeu..... 1222  
15721 CCGCGCGCGCACGCGCGCGCGCGCGCTCGCGCGCTCTTCGCGCC 15770  
1223 LysArgLysSerAlaPheMetAlaProValProThrLysLysArgAsnLe 1239  
15771 CCGCGCGCTTCGGTGGAGCTCGAGACCCAGACCG..... 15809  
1239 uValLeuArgHisGlySerSerSerSerSerSerSerSerSerSerSer 1256  
15810 .....GCGGCTCGCTCGCGC 15825  
1256 LuMetGlyArgArgArgGlyLeuArgValProProProSerSerArgGly 1272  
15826 GGCGCGCGCTCGAGCTCGCGCGCGCGCTCGCGCTCGCGCGACG... 15872  
1273 CysLeuLeuProArgLysProSer...ProProArgAlaMetAlaSer... 1287  
15873 .....CCTGTGCGCGCGCGCTCGACCTCGAGAGTTCGATCCG 15913  
1288 .....LeuProGlnSerSerHisProArgArgProProPheLeu. 1300  
15914 GGTACGCGCTCGCTCGCGCGACCGCGCGCGCGCTCGCGCTCGCG 15963  
1301 .....ProGlnAlaArgLeuSerAlaAlaPheGlnGlyAlaMet 1313  
15964 ACGGCGCTGCTCTCGCGCGCGCGCGCGCTTCACGCGCGCGCTCGT 16013  
1314 LysThrIysValLeuPro.ProArgLysGlyArgGlyLeuLysLeuGluA 1330  
16014 GGTCCGCGCGCGCGCGCGCGCGCTCGCGCGCGCTCGCGCGCG 16050  
1330 laileValGlnLysIleThrSerProSerProSerLysPheAlaCysLys 1346

16051 .....CCT 16053  
1347 AlaProGlyAlaSerProGlyAsnProLeuSerProSerLeuSerAspLy 1363  
16054 GCGCGGGTTCGCGCGCG...CGGCTGCACCGCGCGCTCCACC... 16092  
1363 saspArgGlyLeuLysGlyAlaGlyLysSerProValGlyValGluGlu 1380  
16093 .....GCGCGCACACCCACAGAGATTCCTCATGA 16123  
1380 LysLeuValAsnValGlyThrGlyGlnLysLeu.....ProThrSerGly 1394  
16124 GCGCTAGCGCTCGACACCTCAGAACTCTCGCGAGACCGCGCGG 16173  
1395 AlaAspProLeuCysArgAsnPro.....ThrAsnArgSerLeuLy 1408  
16174 GAGGACACAG...CGTGCACCTCGCGCGAGAACTCGACACCGCTTCG 16219  
1408 sgLysLeuMetAsnSerLysLysLeuSerSerThrAspCysPheLysT 1425  
16220 TGGACTCGGCTACGACTCCCTCGCGCTGTGA..... 16253  
1425 hrGluAlaPheThrSerProGluAlaLeu.....Gln 1435  
16254 ..GAGCGCGCGCTGCTCCAGCAGCGTACGCGCTCGCTGACTGACGA 16301  
1436 ProGlyGlyThrAlaLeuAlaProLysLysArgSerArgLysGlyArg... 1451  
16302 GAGGTCGCGCGCTGGGACCGCGCGACTGCTCGACGAGGTCAACA 16351  
1452 AlaGlyAlaHisGlyLeuSerLysGlyProLeuGluLysArgProTyL 1468  
16352 CCACCGCGCGACCGCTGAGC..... 16373  
1468 euGlyProAlaLeuLeuLeuThrProArgAspArgAlaSerGlyThrGln 1484  
16374 .....GGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16394  
1485 GlyAlaSerGluAspAsnSerGlyGlyGlyLysLysProLysMet... 1500  
16395 GGTCCG 16444  
1501 .....GluLysLeuGlyProAlaSer..... 1507  
16445 GCGGACCTCAGCAGAGTGGCTCTGGTCACTCGCGCGCGCGCGCG 16494  
1508 .....GlnProGluGlyArgProCysGln..... 1516  
16495 GCGCAGTTCGCGCGCTCCCGCACCGAGCGTCCGTCAGACGCAACTGG 16544  
1516 ..... 1516  
16545 GACCGATCAGCGGAGCGCGCGCTCGCGCGGTCACACGACACGCGG 16594  
1517 .....ProGlnThrArgAlaGlnLysGlnProGlyHisThrAsnTyS 1531  
16595 TAGAACTCGCTCTCGAGCGCGCTCTCACCAGCGCGTACGTCAGCAT 16644  
1531 erSerTyrSer.....LysArgLysArgLeuThrArgGly... 1542  
16645 CGAGAACAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16694  
1543 .....ArgAlaLysAsnThrSerSerPro..... 1551  
16695 TGAACCGCGCACCTCTCCGCGATCGCGCGCTCCCGCGCGCGCGCG 16744  
1552 CysLysGlyArgAlaLysArg.....Arg 1561  
16745 GTCCGCGGAGTGGCGGAGACCGCTCCGCGCGGAGAGTGGCGCGAA 16794  
1561 lngGlnValLeuLeuProAlaGluProGluIleArgLeuLys 1577  
16795 CTCTCTCGCGCGTACCGCTCCAGGTGCTGCTGGCGCGCGGTACGACG 16844

PT	/note= "ORF 27; encodes AAB07559"
FT	CDS
PT	/tag= e
FT	/note= "ORF 26; encodes AAB07560"
FT	CDS
PT	/tag= f
FT	/note= "ORF 25; encodes AAB07561"
FT	CDS
PT	/tag= g
FT	/note= "ORF 24; encodes AAB07562"
FT	CDS
PT	/tag= h
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 23; encodes AAB07563"
FT	CDS
PT	/tag= i
FT	/note= "ORF 22; encodes AAB07564"
FT	CDS
PT	/tag= j
FT	/note= "ORF 21; encodes AAB07565"
FT	CDS
PT	/tag= k
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 20; encodes AAB07566"
FT	CDS
PT	/tag= l
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 19; encodes AAB07567"
FT	CDS
PT	/tag= m
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 18; encodes AAB07568"
FT	CDS
PT	/tag= n
FT	/note= "ORF 17; encodes AAB07569"
FT	CDS
PT	/tag= o
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 16; encodes AAB07570"
FT	CDS
PT	/tag= p
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 15; encodes AAB07571"
FT	CDS
PT	/tag= q
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 14; encodes AAB07572"
FT	CDS
PT	/tag= r
FT	/note= "ORF 13; encodes AAB07573"
FT	CDS
PT	/tag= s
FT	/note= "ORF 12; encodes AAB07574"
FT	CDS
PT	/tag= t
FT	/note= "ORF 11; encodes AAB07575"
FT	CDS
PT	/tag= u
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 10; encodes AAB07576"
FT	CDS
PT	/tag= v
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 9; encodes AAB07577"
FT	CDS
PT	/tag= w
FT	/note= "ORF 8; encodes AAB07578"
FT	CDS
PT	/tag= x
FT	/note= "ORF 7; encodes AAB07559"
FT	CDS
PT	/tag= y
FT	/note= "ORF 6; encodes AAB07560"
FT	CDS
PT	/tag= z
FT	/note= "ORF 5; encodes AAB07561"
FT	CDS
PT	/tag= AA
FT	/note= "ORF 4; encodes AAB07562"
FT	CDS
PT	/tag= AB
FT	/note= "ORF 3; encodes AAB07563"
FT	CDS
PT	/tag= AC
FT	/note= "ORF 2; encodes AAB07564"
FT	CDS
PT	/tag= AD
FT	/note= "ORF 1; encodes AAB07565"
FT	CDS
PT	/tag= AE
FT	/note= "ORF 0; encodes AAB07566"
FT	CDS
PT	/tag= AF
FT	/note= "ORF -1; encodes AAB07567"
FT	CDS
PT	/tag= AG
FT	/note= "ORF -2; encodes AAB07568"
FT	CDS
PT	/tag= AH
FT	/note= "ORF -3; encodes AAB07569"
FT	CDS
PT	/tag= AI
FT	/note= "ORF -4; encodes AAB07570"
FT	CDS
PT	/tag= AJ
FT	/note= "ORF -5; encodes AAB07571"
FT	CDS
PT	/tag= AK
FT	/note= "ORF -6; encodes AAB07572"
FT	CDS
PT	/tag= AL
FT	/note= "ORF -7; encodes AAB07573"
FT	CDS
PT	/tag= AM
FT	/note= "ORF -8; encodes AAB07574"
FT	CDS
PT	/tag= AN
FT	/note= "ORF -9; encodes AAB07575"
FT	CDS
PT	/tag= AO
FT	/note= "ORF -10; encodes AAB07576"
FT	CDS
PT	/tag= AP
FT	/note= "ORF -11; encodes AAB07577"
FT	CDS
PT	/tag= AQ
FT	/note= "ORF -12; encodes AAB07578"
FT	CDS
PT	/tag= AR
FT	/note= "ORF -13; encodes AAB07579"
FT	CDS
PT	/tag= AS
FT	/note= "ORF -14; encodes AAB07580"
FT	CDS
PT	/tag= AT
FT	/note= "ORF -15; encodes AAB07581"
FT	CDS
PT	/tag= AU
FT	/note= "ORF -16; encodes AAB07582"
FT	CDS
PT	/tag= AV
FT	/note= "ORF -17; encodes AAB07583"
FT	CDS
PT	/tag= AW
FT	/note= "ORF -18; encodes AAB07584"
FT	CDS
PT	/tag= AX
FT	/note= "ORF -19; encodes AAB07585"
FT	CDS
PT	/tag= AY
FT	/note= "ORF -20; encodes AAB07586"
FT	CDS
PT	/tag= AZ
FT	/note= "ORF -21; encodes AAB07587"
FT	CDS
PT	/tag= BA
FT	/note= "ORF -22; encodes AAB07588"
FT	CDS
PT	/tag= BB
FT	/note= "ORF -23; encodes AAB07589"
FT	CDS
PT	/tag= BC
FT	/note= "ORF -24; encodes AAB07590"
FT	CDS
PT	/tag= BD
FT	/note= "ORF -25; encodes AAB07591"
FT	CDS
PT	/tag= BE
FT	/note= "ORF -26; encodes AAB07592"
FT	CDS
PT	/tag= BF
FT	/note= "ORF -27; encodes AAB07593"
FT	CDS









1350 a..... 1350  
55493 ACCGCCGAGAGTGTGGGCTCTCGAAGAGCCCTTCATGTCGAGGCC 55542  
1350 ..... 1350  
55543 ACCCCGCTCTCAAGCCGAGCGTCCCGCATCAACGCTGACGGGAC 55592  
1351 ..... SerProGlyAsnProLeuSerProSerLe 1360  
55593 GTCCAGAGATCCAGTGGTCAACCGGAGACGCCGCTTCCAGAGCC 55642  
1360 userAspLysAspArgGlyLeuLysGlyValAlaGlySerProValGly 1377  
55643 TCATCCACGC.....CTTCAAGAGCCGPAAGGGAGATCCCAAGGTG 55683  
1377 aGlu.....GluGlyLeuValAsnValGlyThrGlyGln 1388  
55684 CTCACACACAGCTCAACACCAAGGAGCGGATGCGGAGACCCGA 55733  
1389 LysLeuProThrSerGlyAlaAspProLeuGlyAsnPro..... 1402  
55734 GAGCGCCCTGCGACCCCTGCTCGGCTCCGCTCGACACCACTGCTGCTCC 55783  
1403 .....ThrAsnArgSerLeuLysGlyLysLeuMetAsnSerLysL 1416  
55784 CGGGCTCATGTCACGCGCGGACGCGCGCTCATGAGCCGCCCGC 55833  
1416 ys.LeuSerSerThrAspCysPheLysThrGluAlaPheThrSerProG 1432  
55834 GGGGCGAGCGGACCGCGCGCGGCTCGAGCGC...ACTGCGCGCGC 55880  
1432 uAla.....LeuGlnProGlyLysThrAlaLeuAlaProLysLysArg 1447  
55881 ATCTGGGCGGAGACCTCGGCGAGGACAGGTGCGGCCGACGAGGACT 55930  
1447 erArgLysGlyArgAlaGlyAlaHisGlyLeuSerLysGlyProLeu... 1462  
55931 CGCCGCGCTGGGCGGCACTCCATCCAGCGCATCAAGATCAACCAACCGGG 55980  
1463 .....GluLys 1464  
55981 TGGAGAGACTGTCGAGCGGAGCTGTCATCGCGCTCTGTCAGAGC 56030  
1464 sarGProLysLeuGlyProAlaLeuLeuThrProArgAspArgAlas 1481  
56031 CGCACGCTGGCGGCGATGAGACCAAGCTCCAGCGCACGCTCAAGGGGA 56080  
1481 erGlyThrGlnGlyAlaSerGluAspAsnSerGlyGlyGlyLysLys 1497  
56081 GCGGAGAC.....GCTGACACCGGA 56100  
1498 ProLysMetGluGluLeuGlyProAlaSerGlnProGlyLysArgPr 1514  
56101 CCF.....GCCCGGCTGCTGA...CCGATCCGCGGCGC 56132  
1514 ocYsGlnProGlnThrArgAlaGlnLysGlnProGlyLysHisThrAsnTyr 1531  
56133 TGGCGCTCTGTCATGCG...CGAGTCATCTCGACACCTACGCTGG 56179  
1531 erSerTyr..... 1533  
56180 GAGACCACTCGGGCTGTGCGCGAGATCCCGCTCCCTGCGTCAACCT 56229  
1534 .....SerLysArgLysArgLeuTh 1540  
56230 GACCTCGTGGCCACACAGTGGCGGCGCCGACAGCTGCGGTGAAC 56279  
1540 rArgGlyArgAlaLysAsnThr..... 1547  
56280 TCGGCGGCTCGGCGCGACAGGCTGCTCTCGCGAGAGGTGAGAGC 56329

1548 .....ThrSerSerProCysLysGlyArgAlaLys 1557  
56330 CGCGCGCGCGCGCGGCTGCGGAGAGCCCTCGCGCGGAGAGCTGAGAC 56379  
1558 ArgArgArgGlnGlnGlnValLeuPro..... 1566  
56380 CGCGGAGCTCTGTCAGACCCGCGCGGACCGGTCACCAACAGCCGCG 56429  
1567 .....LeuAspProAlaGluProGluLysLeuArgLysTyr 1579  
56430 TCATGGCGAGCGAGAGATGCTGCTCGCCT..... 56460  
1579 LeuSerSerCysLysArgLeuArgSerAspSerArgThrProAlaPhe... 1594  
56461 .....CGACGAGGCGCGGAGAACCCGTTGCCCCGT 56490  
1595 .....SerProPheVal 1598  
56491 GCGGAGGAGCACCGAAGCGGCTGTCGAACGGCGCGCGGCTGCTGC 56540  
1598 ArgValGluLysArgAspAlaPheThrThrLysThrValAlasns 1615  
56541 CGCGCTGCGAGCGCGTATGTCCTCG...ACTACGGGTACGGGCTGGGA 56589  
1615 erProGlyAspAlaPro...LysProHisArgLysProSerSerSerAla 1630  
56590 GCGCGACACCGCGCGGCTGCGCGCACACCGGAGACTGCGCGCTCA 56639  
1631 SerSerSerSerSerSerSerPheSerLeuAspAlaGlyAlaSe 1647  
56640 CCGTGTGCTGCACT..... 56654  
1647 rLeuAlaThrLeuProGlyLysThrLeuGlnProArgProSerLeuP 1664  
56655 .....CGCGCGCGCGCGCGCGCTTCAACCGCGC... 56681  
1664 rLeuSerSerThrMetHisLeuGlyProValLysLysLysAlaLeuSer 1680  
56682 .....TGGGCGCGAGCGCGCTCAAAACCAACCAACG 56711  
1681 ThrSerCysLeuValCysCysLeuGlyGlnAsnProAlaAsn 1694  
56712 CGGAGG.....CGCTGCGCTGCTGAGCGCGCGGAGC 56744



Ratio: 0.647 Gaps: 104  
Percent similarity: 38.107 Percent identity: 22.961

Alignment block:  
US-09-508-821B-6 x US-07-945-283-1/rev ..

Align seg 1/1 to reverse of: US-07-945-283-1 from: 1 to: 8438

```

31 ArgGlnProSerGlnAlaGlyLeuSerCysAspArgGlnArgLeuAla 47
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7236 CGGCGCCATCTGGCGCGCATGCA.....AAGCAGCGGAGAGA 7193
47 aLYAspTYrTYrAsnProGlnProTYrProSerTYrGluGly..... 61
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7192 AGCGGGGAGAAAAATCCATTCGCGCGCGCTCGGGGAAGTCCGCGCGA 7143
62 .....GlyAlaGlyThrProSerGlyThr 69
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7142 AAATCGGCGCATTCGTCCTTACCTGGGGCGGCTCTCTTCGGGGCGCT 7093
70 AlaAlaValAlaAlaAspLYsTYrHisArgGly..... 81
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7092 .....TATAAGCGGCTCTCCATCGTAGCAC 7067
82 .....SerLYsAlaLeuProThrGln..... 88
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7066 TTCACTCGCTGAGTAGGACGACATCGTCTCTGCCAACCGAGGG 7017
88 ..... 88
7016 ATCCGATCGTCTCGCTCGGCGGGGACTGTGAACACTCCGGCTCTCGG 6967
89 .....GlnGlyLeuGlnLYrArgProAlaPheProGly.....Ty 100
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6966 CGGCTACAGCCCTCGACGAGCGCCGACCCACAGAGGCTCTCGGCCAC 6917
100 rGlyValGlnAspSerSerProTYrProGlyArgTYr..... 112
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6916 CAGAGAGAGTCTCTCTCTCTCCGCGCGCTCTCTCTCTCTCTCTCT 6867
113 .....AlaGly, GluGlu..... 116
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6866 GCGCCCGCTCTCGGCGCTCGGCGCGCGCTCTCTCTCTCTCTCTCT 6817
116 ..... 116
6816 CCGATCGAGACCATGCGCGAGACTCTTTGACTTCATCGAGACCGAGG 6767
117 .....SerLeuGlnAlaTrpGlyAlaProGlnProProProGlnPro 131
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6766 CAACCTTCAGCAGCTCTG.....CGGCGCGCGCGCGCGCGCG 6726
132 .....GlnProLeuProAlaGlyValAlaLYsTYrAspG1 143
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6725 AGAGAGAGGCTGCGCTCGGCGCGCGCGCGCGCGCGCGCGCGCG 6684
143 uAsnLeuMetLYsTYrAlaValProProSerArg.....GlnTYrAlaG 159
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6683 .....GCTCCGCGCGCGCGCTCTCTCGGCGAGGATCTCTCTCG 6641
159 luGlnLYAlaGlnValProPheArgThrHisSerLeuHisValGlnGln 175
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6640 CCGGCGCGCTCTCTCTCG.....ACGAGC 6615
176 ProProProGlnGlnProLeuAlaTYrProLYsLeuGlnArgGlnLY 192
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6614 CCGCGAGCGCGAGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 6579
192 sLeuGlnAsnAspIleAlaSerProLeuProPheProGlnLYrHisP 209
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6578 .....CCGAGGCCACGAGCGCG..... 6561
209 heProGlnHisSerGlnSerPheProThrSerSerTYrTYrSerSer 225

```

```

6560 ..... ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
226 ValGlnGlyGlyGlnGlyAlaHisSerTYr..... 236
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6534 GGCAGCGCGCGCGGCTCGGCGAGATCGTGGTGTAGACGACGAGAC 6484
237 LYsSerCysThrAlaProThrAlaGlnProHisAspArgPro..... 250
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6483 GAGGAGGAGGAGAGCGCGGTCGCCGCGCGCGGTCGCCGCGCGC 6434
251 .....LeuThrAlaSerSerSerLeuAlaPro 259
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6438 TCTCCATCAGGCTCGGCGAGCGTACCTCGTCTCGGCGCGCGGTC 6384
260 GlyGlnArgValGlnAsnLeuHisAlaTYrGlnSerGlyArgLeuSerTY 276
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6383 GGGCG.....GCTCCGCGCGCGCGCGCG..... 6360
276 rAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnAlaL 293
6360 ..... 6360
293 euGlnSerArgHisHisAlaGlnGlnThrLeuHisTYrGlnAsnLeuAla 309
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6359 ..CGACCGCGCGACACTCGCAGCG..... 6336
310 LYsTYrGlnHisTYrGlyGlnGln..GlyGlnGlyTYrCysGlnProAsp 325
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6335 .....CAGAGCGCGCGCGCGCGCGCGCGCGCGCTCTCTCTCTCT 6292
326 AlaAlaValArgThrProGlnGlnTYrTYrGlnThrPheSerProSerSe 342
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6291 GCAGCC.....TCCCGCGCGCTC 6275
342 rSerHisSerProAlaArgSerValGlyArgSerProSerTYrSerSer 359
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6274 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6225
359 hrPro.....SerProLeuMet.....ProAsnLeuGluAsnPhe 370
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6224 CCGCGGAGATGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6175
371 ProTYrSerGln.....GlnProLeuSerTh 379
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6174 CCGCGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6125
379 rGlyAlaPheProAla..... 384
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6124 GCGCGAGCGCGCGAGCGCGCGCGCGCGCGCTCTCTCTCTCTCTCTCT 6075
384 ..... 384
6074 CCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6025
385 .....GlyLeuThrAspHisSerHisPheMetProLeu..... 395
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6024 GACCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5975
396 .....LeuAsnProSerProThrAspAlaThrSerSer 407
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5974 GCGCGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5925
407 alaSPThrGlnAlaGlyAsnCysLYsProLeuGlnLYsAspLYsLeuPro 423
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5924 TCTCT..... 5920
424 GluAsnLeuSerAspLeuSerLeuGlnSerLeuThrAlaLeuThrIle 440
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5919 .....TCCAACCCCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5882
440 uGlnValGluAsnIleSerAsnThrValGlnGlnLeuLeuSerLYsA 457
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

5881 CGCCGTCGAGCC..... 5869  
457 IaaIaValProGlnLysGlyValLysAlaLeuValSerGlyThrPro 473  
5868 .....GTCCGCCCGCCGCC 5854  
474 GluGlnHisLysSerGlnHisCysSerProGlnGlySerGlyTyrSerAl 490  
5853 GAGAGAGAGAG.....GACGACTCGGCTCGCCGG 5822  
490 aglProAlaGlyThrProLeuSerGlnProProSerSerThrProGlns 507  
5821 GGAC...GGCGGGCGCCCTCGACAGAGAGAGCCCGCCGCGC..... 5782  
507 eTrHrHisAlaGlnProGlnLysAlaPTrLysSerGlySerGlnAsp 523  
5781 .....CGCGCGGAGAGAGC 5767  
524 ProLeuGlnArgSerPheLeuTyrCysAsnGlnAlaArgly..... 537  
5766 GCCTTCGCGCGCGCGCGGCTTCAGACGCTCCAGACAGCGCGCGAGCGA 5717  
538 .....SerProAlaArgValAsnSerAsnSerLysAla..... 548  
5716 CTCGACCTCTCCCGCCGCTCGCCCTCGCCCGCCCGCCCGCCGCG 5667  
549 .....LysProGlnSerValSerThrCysSerVal 558  
5666 CGCCGCGCGCGCGCGCGCGGCTCGGCTCTCTCTCTCTCTCTCTCTCT 5617  
559 ThrSerProAspAspMetSerThrLysSerAspAspSerPheGlnSerLe 575  
5616 TCGTCT 5567  
575 uHisGlySerLeuProLeuAspSerPheSerLysPheValAlaGlyGlnA 592  
5566 CGGCGCGCGCGCGCCACCTCGCCGC.....GCCGGGCGCGC 5532  
592 rGAspCysProArgLeuLeuLeuSerAlaLeuAlaGlnGlnAlaSerAla 608  
5531 CACGCTCTCGCGG.....GGCGCGCGCGCGCGCGCGCGCTCGGCGC 5488  
609 SerGlnLeuLeuGlyLeuGlnGlnAlaAlaGlyGlnLysAlaSerAla 625  
5487 TCTCTC.....GCCTCGCGACCTCTCTCTCTCTCTCTCTCTCTCTCT 5450  
625 aTrpAlaGlnAlaProSerLeuValLysAspSerSerLysProProPhe 642  
5449 CGCGCGCGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGAGAGCGCGCT 5400  
642 eTrLeuGlnAsnHisSerAlaCysLeuAspSerValAlaLysSerAlaTrp 658  
5399 CCACCAACACACAC.....CTCTCGCTCAGCGCGAGCGCGCGCGCGC 5359  
659 ProArgProGlyGlu.....ProGlnL 666  
5358 CGGACCGAGCGCGCGCTCTCACCGCGCTCGGGAGCGCTCGCGCGCGC 5309  
666 aLeuProAspSerLeuGlnLeuAspLysGlyValAsnAlaLysAspPhe 683  
5308 CGACCTCTCGCGCGCGCGCGCGCTACGCGCGCGCGCGCGCGCGCGCTCC 5259  
683 eTrProGlyLeuPheGlnAspProSerValAlaPheAlaThr..ProAspPr 699  
5258 GCGAGGGGCTCTGGGAGAGAGAGAGCTCGCGCGCGCGCGCGCGCTAC 5209  
699 OlyLysLysThrGlnGlyProLeu.....SerPheGlyThrL 711  
5208 CGCGCGCGCGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5159  
711 ys.....ProThrLeuGlyValPro 717  
5158 GAGGAGACAGACAGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5109

718 Ala...ProAspProThrThrAlaAlaPheAspCysPheProAspThr 733  
5108 GCGAGCGCATGCTCGCTCGAGACCGCGAGAGAGCGCGCGCGCGAGACAG 5059  
733 rAlaAlaSerSerAlaAspSerAla..AspProPheAlaTrpProGln 749  
5058 CGCTTCACACACTCTCGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 5012  
750 AsnLeu.....GlyAspAlaCysProArgTrpPolylLeuHisProG 763  
5011 CGGCT 4962  
763 yGlnLeuThrLysGlyLeuGlnGlnGlyLysAlaSerAspGlyTyr 780  
4961 AGCGATGGCGCGCGAGACCGCTCTGAGCGCTCTGAGCGCGCGCGAGC 4912  
780 eTrLysGly.....AspThrHisGln 786  
4911 GCGGTGGCATGAGCGCGCGCTACAGATCGCACAGAGAGCTTCATCT 4862  
787 AlaSerAlaCysLeuGlyPheGlnGlnGlnAspProProGlyGlnLysVa 803  
4861 GCA.....GAGCTTCGCGCGCGCGCTACGCGC 4836  
803 lAlaSerLeuProGlyAspPheLys..GlnGlnGlnValGlyGlyValLys 819  
4835 ACATGGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGC 4786  
820 GlnGlnAlaGlyLysLeuGlnCysProGlnVal..... 831  
4785 GTCGAGGC.....CTCTGCGCGCGCGCGCGCGCGCGCGCTCTGCGC 4748  
832 ...AlaLysAlaAspArgTrpLeuGlnLysSerArgHisCysCysSer 847  
4747 GCGCGCGAGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCTGCGC 4658  
847 hTrAlaAspPheGlyAspLeuProLeuLeuProProThrSerArgLysGln 863  
4697 TGGCTGCGCGCGCGCGCTGCGAGCGCGCTGCTGCGC..... 4663  
864 AspLeuGlnAlaGlnGlnGlnLysSerSerLysCysGlnLeu..LeuG 879  
4662 .....TCCCGCGCTCGCGGTGCGC 4646  
879 ySerProGlnGlnArgProGlyMetGlnAspProLeuSerProLysAlaP 896  
4645 CG 4612  
896 rLeuLeuLysThrLysGlnGlnValGlnGlnValLeu.....Asp 909  
4611 .....GCCTGCTCGAGAGGTGAGCGCGCGCGCTGCTGCGCTGCGCGAGC 4567  
910 SerLysAlaGlyTyrPheLys...SerProCysHisLeuSerGlyGlnSerVa 925  
4565 GCGATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGCTCT 4517  
925 lLeuLeuGlnLysProThrVal.....G 933  
4516 GCGCTCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4467  
933 lYThrGlnSerLysValGlnSerTrpPheGlnLysSerLeu..SerHisWe 949  
4466 GGGCG 4417  
949 tLysProGlyGlnGlnGlnLysProAspGlyGlnArgLaro..... 962  
4416 CGGCGCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4367  
963 .....GlyAspSerThrThr 967  
4366 GCGCGCGAGCG 4317

[illegible]

```
seq_name: /cn2_6/pilotdata/2/ina/5A_COMB.seq:US-08-258-261B-6
seq_documentation_block:
: Sequence 6, Application US/08258261B
: Patent No. 5638949
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
```

```

62 GYLAIAI+VTThProSeGcGYLTHrAlaIalAIAIaIe.....AI 75
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7609 GGCGGGAGACAGCAGGGCGCTACGGGGCCGCAAGCCCTTCACAGCC 7655
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
775 AAsPluSYrTHsAgIyGeIyALeALeUrPrOThInGLnI..... 89
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7655 CCGTGCAGGACATCGGGCGCCCGGTGAATTACAGGAGCGTCGGTGCCCT 7708
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
90 ..GIyLeuGlnGly ArgProAlaPheProGlyTYrGlyValGlnasPe 105
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7709 GGGGCGCGTGGGGCGGGCGGGCGAGGCCAACGTCAGGGCGAGCGCAC 7758
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
105 rSer.....ProtylProGlyArgTyrlAlaIyGlu 116
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

[illegible]



```

631 erLeuValLysAspSerLysProPheSerLeu..... 643
644 .....GluAsnHisSerAlaCysLeuAspSerVal 653
9201 GCTCCAGACGACCCGATCACCAGGCTGACGCTCCAGTGC 9250
9251 ATGGACCGGACCTCTGGAGAGCCCATGAGGTG..... 9289
653 LAlaLysSerAlaTrpProArgProGlyGluProGluAlaLeuProAsp 670
9290 .....AAGCCTGGCCG...CCGCTACGCGGACGAGAGCCG... 9325
670 erLeuGlnLeuAspLysGlyValAsnAlaLysAspPheSerProGlyLeu 686
9326 .....CTGAAAGCCTCTCTCTCTGCGG 9349
687 PheGluAspProSerValAlaPheAlaThrProAspProLysTrpThr 703
9350 CGCTCAACCAACATCGGCATCTCGAGCGCGCTCGGCGCTGCGAGG 9399
703 rglYProLeuSerPheGlyThrLysProThrLeuGlyValProAlaPro 720
9400 GTGCGCAAGATCG.....TCGCTCCCTCGCCCATGACGCGCTGCGCC 9443
720 sPrProThrThrAlaAlaPheAspCysPhePro.....AspThr 732
9444 CACCTCCACACGCGCGCGCAT...CCCTGATGATGAGATCA 9489
733 ThrAlaAlaSerSerAlaAspSerAlaAsnProPheAlaTrpProGlu 749
9490 CTGCGCATCGACGCTGTATAC...CCGAGCTGTGGCGCGCA 9533
749 uAsnLeuGlyAspAlaCysProArgTrpGlyLeuHisProGlyGluLeu 766
9534 CGAA.....GATAGACTCCCGC..... 9552
766 hrLysGlyLeuGlnGlyValLysAlaSerAspGlyIleSerLys 782
9553 .....CGCGCGGCGTCTCGGCTTCGAGACCTCCGCGCAC 9588
783 AspThrHis.....GluAlaSerAlaCysLeuGlyPheGln 795
9589 AAGCCCACTGATCATCTCGAGGAGCTCCGCGCCCTG.....TC 9629
795 uGluAspProProGlyGluLysValAlaSerLeuProGlyAspPheLys 812
9630 GGGCGAGCGCGCCACCTCACAGAGCGCGCTCGACGCTCCCGCGGCT 9679
812 InGlnGluValGlyValLysGlnGluAlaGlyLysLeuGlnCys 828
9680 GTGCGCTGCTCTCTGCGGCAAGAGAGCGCGCGCTCCGCGCGCACG 9729
829 ProGlu.....ValAlaLysAlaAspArgTrpLeuGln 840
9730 AAGCGGCTCCGACACATCTCTCGCCACAGACCTCCCTTATGGA 9779
840 pSerArgHisCysCysSerThr.....AlaAspPheGlyAspLeuPro 854
9780 TGTGGCTATTCGACGAGCGACACCGCGCCACTGACACCGCGCG 9839
855 LeuLeuProProThrSerArgLysGluAspLeuGlnAlaGlnGluTrp 871
9830 CTCTCTGGCGCGCGACCGGACG..... 9853
871 rSerSerLeuCysGlnLeuLeuGlySerProGlnGlnArgProGlyMet 888
9854 AGCTCTCTCTCGCGCTGACTGCTCGCCACAGACGCG..... 9895
888 LAspProLeuSerProLysAlaProLeuIleCysThrLysGlnGluVal 904
9896 .....CCCGAGACACGCTTCTCGGCGGAGAGCGAGAGCGAG 9931
905 GlnGluValLeuAspSer.....LysAla.....GlyTrp 914

```

```

9932 GCAAGGTCTGCTCTCTCTCTCTGAGGACAGGCTCGACAGGAGATG 9981
994 pGlySerProCysHisLeuSerLysGlnSerValIleLeuLeuGlyPro 9931
9982 GCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10031
9931 hrValGlyThrGlnSerLysValGlnSerTrpPheGlnSerSerLeuSer 947
10032 ATGGAGAGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10060
948 HisMetLysProGlyGlnGluProAsp.....GlyGluArgAlaPro 962
10061 .....AGTGAAGCTGCTGCGCTCTCTCTCTCTCTCTCTCTCTCT 10101
962 oGlyAspSerThrThrSerAspAla..... 970
10102 CCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10151
971 .....SerLeuAlaGlnLysProAsnLysProAlaVal 981
10152 GGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10201
982 ProGluAlaProIleAlaLysLysGluProValProArgGlyLysSer 998
10202 TCGTGGGCGCACAGCGGCGGAGATGCGCGCGCTCTCTCTCTCTCT 10251
998 uArgSerArgArgValHisArgGlyLeuProGluAlaGluAspSerPro 1015
10252 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10286
1015 ys.....ArgAlaProValLeuProLysAspLeuLeuProGlu 1028
10287 GCGGAGAGAAAGCGCTCACACCGCTCGCGGACAGCGGCGGATGCG 10336
1029 SerCysThrGlyProProGlnGlyGln.....MetGluValAla.. 1041
10337 TCGAGCTCGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10386
1042 .....GlyAlaProGlyValArgValAlaSerGln 1051
10387 AGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10436
1051 LysLeuProArgMetCysThrArgSerLeuThrAlaLeuSerGlnPro 1067
10437 CGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10486
1068 ...ThrProGlyProProGlyLeuThrThrThrProAlaPro..... 1080
10487 TGTTCGCGCGGAGATCCGCTCGATCACCTCTCTCTCTCTCTCTCT 10536
1081 .....ProAspLysLeuGly 1086
10537 GACGCGCTCCACAGAGCTCGCGGAGGCTCTCTCTCTCTCTCTCTCT 10586
1086 LysValArgAlaAlaPheLysSerLysArgValGlyLysProSer 1102
10587 GACGTGCGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10618
1103 ProLysAlaAlaSer...SerProSerAsnProAla..... 1113
10619 CGGCGACCGAGGCTCGAGGCTCGAGGCTCGAGGCTCGAGGCTCTCT 10668
1114 .....AlaLeuProValAlaSerAspSerProMetGlySerLysTr 1128
10669 AACCTCGGCAAAACGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10718
1128 hrLysGluThrAspSerProser..... 1135
10719 CGATGGGAGATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10768
1136 ThrProGlyLysAspGlnArgSerMetIleLeuArgSerArgThrLys 1152

```

10769 TCAGCCCTCCGAGAGAGCTCGAGCGCTACCGCTCGATCCGCTGCTC 10818  
1152 hrGlnGlnIlePheHisSerLysArgArgProSerGluGlyArgLeu 1168  
10819 GG.....CTCAATCGAGAGAGAGCCACTCG..... 10849  
1169 ProAsnGys.....ArgAl 1173  
10850 CCGCGCTGCTCTCTCTCGGGGAGCTCTCTACCGAGGCGCTCGCGCTC 10899  
1173 arThrLysLeuLeuAspAsnSerHisLeuProAlaThrPheLysValS 1190  
10900 GACTGGAGAGACTTCTTCGCGGCCCTACGCTCCCGCGAAGGTCCCTCC 10949  
1190 erSerSerProGlnLysGlu.....GlyArgValSerGlnArgAla 1203  
10950 CACCTACCCCTTCAGGAGAGCGGTTCGTGCTCGACGTCTCCAGGAGC 10999  
1204 ArgVal..... 1205  
11000 AAGCGTTCGAGCTCGCTCCGAGGCGCTGACCTCGCGCGACCAATCCG 11049  
1206 .....ProLysProGlyAla..... 1210  
11050 CTCTCGCGCGCGCGCTTCGCGCTTCGCGACCGCGGTGGCTTCTTTAC 11099  
1211 .....GlySerLysLeuSerAspArg 1217  
11100 AGCGGCGCTCTCCCTCGAGAGACCCGCTGCTCGAGGCGCATCGCTCT 11149  
1218 ProLeuHisAlaLeuLysArgLysSerAlaPhe..... 1228  
11150 TCGGCACACCATCTACCGCGCACCGCTTCTCGAGCTCGCCCTGCAC 11199  
1229 .....MetAlaProValProThrLys.....LysArgA 1238  
11200 GTGCGCCACCGCGCTCGCGCTTCGACACCTCGAGAGCTCACCTCGAGC 11249  
1238 snLeuValLeuArgHisGlySerSerSerSerSerAlaSerAlaMet 1254  
11250 CCCTCTCGCTCTCCCATCGAGGACACCGTCTCTCTCCAGATCTCGTCG 11299  
1255 Gly..... 1255  
11300 GGCCCTGGAGCGAGCAGCAGAGAGGCGCTCTCTTTCCATAGCCGACAA 11349  
1256 .....GluMetG 1258  
11350 GAGGAGCGCTTCAGGATGGCCCTGGACTCGCCACGCCAGCGGCTCTCT 11399  
1258 lyArgArgArgGlyLeuArgValProProProSerSerArgGlyCysLeu 1274  
11400 CTCGCGCGGCGCCCATCTCTCCGCGGATCTCCAGAGTGC..... 11443  
1275 LeuProArg.....LysProSe 1280  
11444 CTCCTCTGAGTCCATCCGCTGGACCTCGAAGGCTCTACGCAACCTC 11493  
1280 rPro.....ProArgAlaMetAlaSerLeuProG 1290  
11494 GCACACTCGGCTTGCTACGCGCCGAGTTCAGGCGCTCTCGCTCGCT 11543  
1290 InSerSer.....His 1293  
11544 CTACAAGCGCGGAGAGCTCTTTGCGAAGCCAAAGCTCCCGAAGCGG 11593  
1294 ProArgArg.....ProProPheLeuProGlnAlaArgLeuSerAlaAlaP 1309  
11594 CCGAAAGGATGCCCGCGGTTCGCC...TCCACCTCGGCTGTGCTGAC 11640  
1309 eGlnGlyAlaMetLysThrLysValLeuProProArgLysGlyArgGlyL 1326  
11641 AGCCCTCTGATGATCGCTGGCTTTGAGGACGAGGAGGAGGCTGCG 11690

1326 eu..... 1326  
11691 TCTGCCCTTCTCGTGGAGGAGTCTCGTGGCTCGCTCGGTGGCCACCA 11740  
1327 .....LysLeuGluAlaIleValGlnLysIleThrSerProSer..... 1339  
11741 CTTTGGCGTGGCTTCCACCGTCCCAAGGTGAATCTCGCTCTGATC 11790  
1340 .....LeuLysLysPheA 1344  
11791 GTCCTGGCGAGCGCGAGGTGACCTCTTCTGCTCGGTGCAAGCGTGC 11840  
1344 lacys..... 1345  
11841 GATCGGAGAGAGTCCGCGCGAGCTCCGACACCGCGAGCTTCCACC 11890  
1346 .....LysAlaPro..... 1348  
11891 ATGATCGGCTTCTCGCTCGTCACTGGAGGAGCTCCAAAGCCCACTTCA 11940  
1348 ..... 1348  
11841 CCGCTCGCGCGCGCGAGGCGCTCTTCTCGGCACAGCGGCGCACATCT 11990  
1349 .....GlyAlaSerProGlyAsnProLeuSer...ProSerLeu 1360  
11991 CGCGCTCGAGCGCGCGCTCGCGCGCTACCGAGCTCGCTCGCTCCGAA 12040  
1361 SerAspLysAspArgGlyLeuLys...GlyAlaGlyGlySerProValG 1376  
12044 GCGCGCTCGACAGCGCGCTTTCGCTCCGCGCGCTCGCTCGCGCC..... 12085  
1376 yValGlnGlnGlyLeuValAsnValGlyThrGlnLysLeuProThr. 1392  
12086 .....CCTTCATGATCGACCGCGAGGCGCTCTCGCGAGCGCGCCACGA 12131  
1393 .....SerGlyAlaAspProLeuCysArg.....AsnProThrAsn 1404  
12132 GCGCACCGCGCTCGACCTCGCTCTTGAAGCTGGCTCGCGAGCAAC 12181  
1405 ArgSerLeuLysGlyLysLeuMetAsnSerLysLysLeuSerSerThrAs 1421  
12182 GCCTCGCTCTCGCGCTCTCTCTCGACCGAGCGCGCTCGCGCAC 12231  
1421 pCysPheLysThrGluAlaPheThrSerProGluAlaLeuGlnProGlyS 1438  
12232 CACACCGAGAGCGCTCAAGACTCG.....CTCAGCGCGCTCTG 12275  
1438 lyThrAlaLeuAlaProLysLysArgSerArgLysGlyArgAlaGlyAla 1454  
12276 GGGCTCGCGCTCGCGCAAGTGAGCACCCAGACCTCCGCTCTTCC 12325  
1455 HisGlyLeuSerLysGlyProLeuGluLysArgProTyrLeuGlyProAl 1471  
12326 TCGTCGACATCG.....ACCTCAGAGGCGCTCCACAGCGCGCTG 12366  
1471 aleuLeuLeuThrProArgAspArgAlaSerGlyThrGlnGlyAlaSerG 1488  
12367 CTAGGCGCTCGACAGAGAGAGCGGCTCGCTCCGCTCCGCAAGGAA 12416  
1488 luAspAsnSerGlyGly..... 1493  
12417 ACCCTCATCCGAGGTGGGCGAACCAAGCTCGAGCGAGCGCTCATCC 12466  
1494 .....GlyGlyLys.....LysProLysMetGluG 1502  
12467 CGCGCGAAGCACCGAGTGGCGCTCCATATTCGACCAAGGACCTTC 12516  
1502 uLeu...GlyProAlaSerGlnProProGluGlyArgProCysGlnProG 1518  
12517 GACGCGCTCGCTCTCGTACGCGCGG.....AGCGCGAGCGCGCT 12560

```

1518 InthrtglaInLvgInProglYh1sthrAntyYrserYrser 1534
      :: |||::: |||::: |||::: |||::: |||
12561 CGACAGCGCCAGTCCGATCGCGCGGACGCGGCTCACTTCC 12610
      :: |||::: |||::: |||::: |||::: |||
1535 LysArgLys.....ArgLeuThrArgLysArgL 1544
      :: |||::: |||::: |||::: |||::: |||
12611 GCATGTGTGACACCTTGAGATGATCCGGGCGCGCGCGCG.. 12658
      :: |||::: |||::: |||::: |||::: |||
1544 aLysasnThrYrserYrserProGyLysGlyArgLysArgLys 1561
      :: |||::: |||::: |||::: |||::: |||
12659 .....TCGAGGCGGAGCGCGCGGATGCTTACGAGT 12692
      :: |||::: |||::: |||::: |||::: |||
1561 InGlnGlnValLeuProLeuAspPro..... 1569
      :: |||::: |||::: |||::: |||::: |||
12693 CGGTCCAGGTGTCTCCGATACACCTAGGCGACGCGGAGGAGT 12742
      :: |||::: |||::: |||::: |||::: |||
1570 ...AlaGluProGluLysArgLysYrLysYrserYrserYr 1585
      :: |||::: |||::: |||::: |||::: |||
12743 TCGGCGGACGCTTGTGTC..... 12760
      :: |||::: |||::: |||::: |||::: |||
1585 uArgSerAspSerArgThrProAlaPheSerProPheValArgVal 1602
      :: |||::: |||::: |||::: |||::: |||
12761 .CCAGCGCATCGCGGACGCGCGGATGATC..... 12790
      :: |||::: |||::: |||::: |||::: |||
1602 YsArgspAlaPheThrThrLecYsThrValAsnSerProGlyAsp 1618
      :: |||::: |||::: |||::: |||::: |||
12791 .....GCCCATCCCGACGCGCTGTCT 12814
      :: |||::: |||::: |||::: |||::: |||
1619 AlaProLysProHisArgLysProSer..... 1628
      :: |||::: |||::: |||::: |||::: |||
12815 TCGCCCAAGCCGACAGGTCCCATCATCTACACCGCTACTATGA 12864
      :: |||::: |||::: |||::: |||::: |||
1629 .....SerAlaSerSer..... 1633
      :: |||::: |||::: |||::: |||::: |||
12865 CTGTCGATCTCGGACATCGAAGCCATCAACGTCCTCATCATCAGC 12914
      :: |||::: |||::: |||::: |||::: |||
1634 .....SerSerSerSerPheSerLeuAspAlaLys 1645
      :: |||::: |||::: |||::: |||::: |||
12915 GCCCGCGCGCGCGCTCGGACGCGCGCGCTTCACTCGACGCGCTCG 12964
      :: |||::: |||::: |||::: |||::: |||
1646 AlaSerLeuAlaThrLeuPro.....GlyGlySerLysLeuGlnPr 1659
      :: |||::: |||::: |||::: |||::: |||
12965 GCGCGAGGTCTTGGCCACGCGGATCGAGGAGTGAAGCGCTCGGC 13014
      :: |||::: |||::: |||::: |||::: |||
1659 cArgProSerLeuProLeuSerSerThrMetHisLeuGly...ProValY 1675
      :: |||::: |||::: |||::: |||::: |||
13015 GCGCTCG.....GCTTCGACGATCGCACCTCGCTCTCGACG 13052
      :: |||::: |||::: |||::: |||::: |||
1675 alserYsAla...LeuSerThrSerYs 1683
      :: |||::: |||::: |||::: |||::: |||
13053 TGACCTGGGCTTCGACGACGACTCTCTCGC 13081
      :: |||::: |||::: |||::: |||::: |||

```

seq\_name: /cgn2\_6/plodata/2/lna/5A\_COMB.seq:US-08-456-837-6

seq\_documentation\_block:

```

: Sequence 6, Application US/08456837
: Patent No. 3643774
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James M.
: APPLICANT: Beck, James Joseph
: APPLICANT: Hill, Dwight Steven
: APPLICANT: Ryals, John Andrew
: APPLICANT: Gaffney, Thomas Deane
: APPLICANT: Lam, Stephen Ting
: APPLICANT: Hammer, Phillip E.
: APPLICANT: Uknes, Scott Joseph
: TITLE OF INVENTION: Genes for the synthesis of
: NUMBER OF SEQUENCES: 22
: NUMBER OF INVENTION: antipathogenic substances
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ciba-Geigy Corporation
: STREET: 7 Skyline Drive

```

```

CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,837
FILING DATE: 01-JUN-1995
CLASSIFICATION: 435
ERROR APPLICATION DATA:
APPLICATION NUMBER: 08/457,205
FILING DATE: 01-JUN-1995
APPLICATION NUMBER: 08/258,261
FILING DATE: 08-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-456-837-6

```

alignment\_scores:

Quality:	493.00	Length:	2095
Ratio:	0.593	Gaps:	102
Percent Similarity:	39.666	Percent Identity:	22.291

alignment\_block:

US-09-508-821b-6 x US-08-456-837-6 ..

Align seg 1/1 to: US-08-456-837-6 from: 1 to: 28958

```

62 GYAlaGlyThrProSerGlyThrAlaAlaValAla.....Al 75
|||::: |||::: |||::: |||::: |||
7609 GCGGCGGACGACGAGCGGCTACGCGGCGGACGCGGCTCTCGACG 7658
|||::: |||::: |||::: |||::: |||
75 asPlysYrLysArgLysSerLysAlaLeuProThrGln..... 89
|||::: |||::: |||::: |||::: |||
7659 CCTTCGAGACATCGCGGCGGAGCTGATGACACGACGCGGTGCT 7708
|||::: |||::: |||::: |||::: |||
90 ..GlyLeuGlnGly.ArgProAlaPheProGlyYrGlyValGlnAsp 105
|||::: |||::: |||::: |||::: |||
7709 GGGGCGGTGGGCGGCGGCGGATCGCCACGATCAGCGCGCGCCAC 7758
|||::: |||::: |||::: |||::: |||
105 rser.....ProYrProGlyYrGlyValGlnGln 116
|||::: |||::: |||::: |||::: |||
7759 CTCCACACGCGGCTGTGCGGAGATGCGCCCTGCTGCGCTGCGG 7808
|||::: |||::: |||::: |||::: |||
116 LysLeuGlnAlaLysProGlyAlaProGlnProProProGlnPro... 131
|||::: |||::: |||::: |||::: |||
7809 GGTGCGGTGTGTGAGACGACGACGACGACGCGCTGCGCGGACA 7858
|||::: |||::: |||::: |||::: |||
132 .....GlnProLeu...ProAlaGln 137
|||::: |||::: |||::: |||::: |||
7859 TCGAGTGGGCGGCTTGTGCGCTTGTGACGCGCGCTGCGCGCGCC 7906
|||::: |||::: |||::: |||::: |||
137 YValAlaLysYrAspGluAsnLeuMetLysYrThrAlaValProPro 154
|||::: |||::: |||::: |||::: |||

```

[illegible]

437. laLeuThrLeuGlnValGluAsnIleSerAsnThrValGlnGlnLeuLeu 453  
8505 CTACGTCGGCAGTCGCGCAGTCCTCGACAGTGTGACGCTTCTGACCGCT 8554  
454. LeuSerLysAlaAlaValProGlnLysGlyValLysAsnLeuValSer 470  
8555 CCTCTCTTGGCATCAGCCCGCGGGAGCCAAAC .....AC 8589  
470. rArgThrProGluGlnHisLysSerGlnHisLysCysSerProGluGlySerG 487  
8590 CTGACACGCTC .....AGACCGCGCTGCTCC .....TCGA 8618  
487. lyTySerAlaGluProAlaGlyThrProLeuSerGluProPro ..... 501  
8619 ATCTGCTGGCAGGCGCTCGAAGACGCGCGCACTGCTCCCGCACCGCTCA 8668  
504. ....SerSerThrProGlnSerThrHi 509  
8669 AGGATTCCCGCCACCGCGCTTCTGCTGGCATCGCGCGCAGCAANTACGCA 8718  
509. sAlaGlu .....ProGlnGluAla 516  
8719 TTGCGAGAGCGCAGCACCGAAGATTCGAGCGTTATGCGCTTCCAGGCA. 8767  
516. sPtyrLeuSerGlySerGluAspProLeuGluArgSerPheLeuTyrCys 532  
8768 .....CGCGCGGTCTTTGCGCGGGCGGTGGCTCATCA 8803  
533 AsnGlnAlaArgGlySerProAlaArgValAsnSerAsnSerLysAlaLys 549  
8804 CGTGGCGCTCAAGAGCGCG .....CGTCTCGGTGCGACCGCTGCTCTCC 8850  
549. sProGluSerValSerThrCysSerValThrSerProAspAspMetSer 566  
8851 TCCTCGCTGCTGCGCTTCCACCTCGCTGCGCCAGCGCTCCGACAGGCGCA 8900  
566 hrLysSerAspSerPheGlnSerLeuHisLysLysLeuProLysLeuAsp 582  
8901 GTGCAACCTCGCGCTCGCGCGGGCGGTCTCGGTATGCGCTCCCGCGAG 8950  
583 SerPheSerLysPheValAla ..... 589  
8951 GCTTCGCTCTCTTCCGCGCTGCGCGCTTGGCGCGAGCGCGCTCC 9000  
590 .....GlyGluArgAspCysP 595  
9001 AAGACCTTCTCGGCAACGCGAGCGGTACGGAGCGGAGAGGCTCAT 9050  
595 roArgLeuLeuLeuSerAlaLeu .....AlaGlnGluAspLeuAla 608  
9051 CGTCTGTGCGCTCGAGCGCTCGGTGACGCGCTTCGCCCGAGGACACGCG 9100  
609 SerGluIleLeuGlyLeuGlnGlnAlaIleGlyGluLysAlaAspLysAl 625  
9101 TCTCGCTCTGTCGCGGCGACCGCATCAACGACGCGCGGTGCGAGC 9150  
625 a .....TrpAlaGluAlaPro 631  
9151 GGTATCAGCGCGCCCAACAGCGCACCTCCACGACAGAGGTCTCGCGCGCGC 9200  
631 erLeuValLysAspSerSerLysProProPheSerLeu ..... 643  
9201 GCTCCACGACGCGCATCACCCTCGCGCGAGCTGCGACGTGCTGAGTGGCC 9250  
644 .....GluAsnHisSerAlaCysLeuAspSerVal 653  
9251 ATGGCAGCGGCACTCTCTGGGAGACCCCATCGAGGTGC ..... 9289  
653 lAlaLysSerAlaTrpProArgProGlyGluProGluAlaLeuProAsp 670  
9290 .....AAGCCCTGGCG .....CGGTCTAGCGCAGCGACGACCG ..... 9325

10661 ..... AGTGAAGCTCTGTCGCCGTCCCTCGGGCCGCACGAAAGGCGCC 10101  
||| :||||| :||| |||||  
962 oglyaspSerThrThrsSerAspa..... 970  
||| :||| :|||  
10102 CCTCCTCTGACCGCGTAGCGTGATCAAGCCGCCCTTTGCCGCTCAT 10151  
||| :||| :|||  
971 ..... SerLeuAlaGlnLysProAsnLysProAlaVal 981  
||| :||||| :||| :|||  
10152 GAGTCCCTGGCGCGCCCTCTGGGGTCTGCTGGCGGGGTGAGCCGCCGCCG 10201  
||| :||| :||| :|||  
982 ProGluAlaProIleAlaLysGlnLysGluProValProArgLysSerLe 998  
||| :||| :||| :|||  
10202 TCGTCGGCACAGCCAGAGSGAGATGCCGCCGCTTCTGTCCAGGCGCT 10251  
||| :||| :||| :|||  
998 uArgSerAArgValHisArgLysLeuProGluAlaGlnAspSerProc 1015  
||| :||||| :||| :|||  
10252 CTCTCCCTGAGAAGAGCGGCGCATGCGG..... CCGT 10288  
||| :||| :||| :|||  
1015 ys..... ArgAlaProValLeuProLysAspleuLeuProGlu 1028  
||| :||| :||| :|||  
10287 GCAGAGAAAGGCGTACCACCGTGGCGGCAAGCGCGCATGCGCGCG 10336  
||| :||| :||| :|||  
1029 SecCysThrGlyProProGlnGlyGln..... MetGlnGlyAla.. 1041  
||| :||||| :||| :|||  
10337 TCAGAGCTGGGCGCTCCGACCTCCAGACTTACCTCGCTCCGCGGGGAGC 10388  
||| :||||| :||| :|||  
1042 ..... GlyAlaProGlyArgGlyAlaSerGln 1051  
||| :||||| :||| :|||  
10387 AGGCTGTCCAGCGCGCGCGTCAACAGCCCGAGGCTACCGCTGATCCGG 10436  
||| :||| :||| :|||  
1051 LysLeuProArgMetCysThrArgSerLeuThrAlaLeuSerLnuProArg 1067  
||| :||| :||| :|||  
10437 CGAGCCCGCGCGCGTGCAGCGCGTGCCTGACGCTTCACCGCCACCAAG 10486  
||| :||| :||| :|||  
1068 ... ThrProGlyProProGlyLysLeuThrThrThrpAlaPro..... 1080  
||| :||||| :||| :|||  
10487 TGTTCCGCGGCAAGTCCGCGTGAGTACGCTCCCTCCAGCCCAAGTG 10536  
||| :||||| :||| :|||  
1081 ..... ProAspLysLeuGln 1086  
||| :||| :||| :|||  
10537 GAGCGCGTCCAGAGAGAGCTCGCGGAGGTCTACCAACAATGCTCTCG 10588  
||| :||| :||| :|||  
1086 LysLysGlnArgAlaAlaPheLysSerGlyLysArgValGlyLysProser 1102  
||| :||| :||| :|||  
10587 GACCTGGAGCTCCCTCTTATT..... CGAACCTGTA 10618  
||| :||| :||| :|||  
1103 PolyAlaAlaSer... SerProSerAsnProLa..... 1113  
||| :||||| :||| :|||  
10619 CGGCGACACAGGCTCAGAGCTCCGACCTCGAGCGCGTACGTAGTATGA 10666  
||| :||||| :||| :|||  
1114 ..... AlaLeuProValAlaSerAspSerProMetGlySerLysT 1128  
||| :||| :||| :|||  
10659 AACCTCCGGCAACCGTCTGCTTCTGAGCGCAGCGAGGGGCTCTCTGA 10712  
||| :||| :||| :|||  
1133 hTrpLysGluThrAspSerProSer..... 1135  
||| :||||| :||| :|||  
10719 CGATGGGCAATCGCTCTCGCTCAGAGGTACGCCCATCCGCTGCTACGC 10766  
||| :||||| :||| :|||  
1136 ThrProGly... LysAspGlnArgSerMetLeuArgSerArgThrLysT 1152  
||| :||| :||| :|||  
10759 TCGCCCTCGGAGAGCTCGGAGCGCTCAACGCTGATCCGCTGCTCTC 10811  
||| :||| :||| :|||  
1152 hArgGlnLysPheHisSerLysArgArgArgProSerGlnGlyArgLeu 1168  
||| :||||| :||| :|||  
10819 GG..... CTCATTCCAGCAAGAAAGGCCACTCG..... 10844  
||| :||| :||| :|||  
1169 ProAsnCys..... ArgAl 1173  
||| :||| :||| :|||  
10850 CCGCGCGTCTCCTCTCCGGGCGGAGCTCTTACCCGAGCGCTCGCGCTC 10899  
||| :||| :||| :|||  
1173 aThrLysLysLeuLeuAspAsnSerHisLeuProAlaThrPheLysValS 1190  
||| :||| :||| :|||

10900 GACTGGAAAGGACTCTTCGTCCGGCCTTCAGCTCCGCCGAAGAAGTCTCCCTCCC 10944

1190 erSerSerProInLysglu.....GlyArgValserGlnArgAla 1203  
:|||||:|||||  
||| :||| |||:

10950 CACCTACCGCTTCACGAGAGCGGTCTCGCTCGACGCTCTCCACGGAGC 10999  
:|||||:|||||  
||| :||| |||:

1204 ArgVal..... 1205  
:|||||

11000 AAGCTTCGACGCTGCCTTCGACAGGCTGACCTGGCCGACCAATCCGG 11049

1206 .....ProLysProGlyAla..... 1210  
|||||:|||||  
||| :||| |||:

11050 CTGCTCGGGCGCGCGCFGCGCTTCGCCGACGCGGTGCTTCTCTTTAC 11099  
:|||||:|||||  
||| :||| |||:

1211 .....GlySerLysLeuSerAspArg 1217  
|||||  
||| :||| |||:

11100 AGSGGGCTCTCCCTCGCAGACACCCTGGCTCGAAGCGCATGCCGTCT 11149

1218 ProLeuHisAlaLeuLysArgLysSerAlaphe..... 1228  
||| :|||||:|||||  
||| :||| |||:

11150 TCGGCACACCATCTACCGGSCACCGCTTCTCGAGCTGGCCCTGCAC 11199

1229 .....MetAlaProValProThrLys.....LysArgA 1238  
||| :|||||:|||||  
||| :||| |||:

11200 GTGCCGCCACCGCTCGCGCTCGACACGCTCGAAGCTCACGTCCGAGC 11249

1238 snLeuValLeuArgHisGlySerSerSerSerSerAsnAlaSerAlaMet 1254  
||| :|||||:|||||  
||| :||| |||:

11250 CCTCTCGCTCTCCATCCAGGACACGCTCTCTCCAGNATCCGTGCG 11299

1255 Gly..... 1255  
|||

11300 GGCGCTGGACGCGCAGGACGAAGGCGCTCTCTTTCATCGCGACAA 11349

1256 .....GluMetG 1258  
||| :|||

11350 GAGGACGCGCTTCAGATGGCCCTTGAGCTCGCCACGCGAGCGCTCTCT 11399

1258 lVArgArgArgLysLeuArgValProProProSerSerArgGlyCysLeu 1274  
:|||||:|||||  
||| :||| |||:

11400 CTGCGGGGACCCATCCCTCTCGCGCATCTCCACGATGTCGCTGCG 11443

1275 LeuProArg.....LysProSe 1280  
|||||  
||| :|||

11444 CTCCTTCGAGTGCCATCCGCTGGACCTCGAAGGCTCTACGCAACCTC 11493

1280 rPro.....ProArgAlaMetAlaSerLeuProG 1290  
||| :|||  
||| :||| |||:

11494 GCCAACCTCGGGCTTCCTTCAGCGCCCGAGTTCACGAGGCTCGCTCGT 11543

1290 InserSer.....His 1293  
:|||||

11544 CTCACGCGGGGAGAGCTCTTTGGCAGCGCAAGCTCCCGGAAGCGG 11593

1294 ProArg...ProProPheLeuProGlnAlaArgLeuSerAlaAlaph 1309  
:|||||:|||||  
||| :||| |||:

11594 CGAAAAGATGCGCGCGGTGCG....TCCACCTGGCTGCTCGAC 11640

1309 eGlnGlyAlaMetLysThrLysValLeuProProArgLysGlyArgGlyL 1326  
:|||||  
||| :||| |||:

11641 AGCGCCCTGCATGCATCGCTTCGCTTAGAGACGAGACAGAGGAGCGTGC 11690

1326 eu..... 1326  
|||

11691 TCTGCCCTTCGTGAGCGAGTCTCGCTGGCTCGGTGCGCACCA 11740

1327 .....LysLeuGluAlaIleValGlnLysIleThrSerProSer.... 1339  
:|||||:|||||  
||| :||| |||:

11741 CTTTGGCGTGGCTTCCACCGTCCACAGGGGTATCTCTCGTCTCGATC 11790

1340 .....LeuLysLysPheA 1344  
:|||||

11791 GTCTGGCCGACGCGCAGGTGACCTTCTTGCTCGGTGACAGCGTGC 11840

1344 laCys..... 1345  
|||  
11841 CATCGGAGACAGCTCCGCGCGCAGCTCCGACCCCGGACGCTCCGAC 11890  
1346 .....LysalaPro..... 1348  
11891 ATGATGGCTCTTCCGCTGCACTGGAGCAGGTCGAAGAGCCCACTTCA 11940  
1348 ..... 1348  
11941 CGGCTCGCCCGGAGCGGGCTCTTCTCGGCACAGCGCGCACGATCT 11990  
1349 .....GlyAlaSerProGlyAsnProLeuSer...ProSerLeu 1360  
|||||..... |||  
11991 GCGGCTGAGCGCCCGCTCGCCGCTACGCGACCTCGCTGCCCTCCGAA 12040  
1361 SerAspLysaspargGlyLeuLys...GlyAlaGlyGlySerProValG 1376  
: : : : : ||| : : : : :  
12041 GCGCCTCGACACGGCGCTTGGCTCCCGGCTCGTGTGCGCC..... 12085  
1376 yValGluGluLeuValAsnValGlyThrGlyGlnLysLeuProThr. 1392  
: : : : : : : : : : : |||  
12086 ...CTTTCATCATGATCGACGCGCAGCGAGCTCGTCCGAGCGCCACGA 12131  
1393 .....SerGlyAlaAspProLeuLysArg.....AsnProThrAsn 1404  
: : : : : ||| : : : : :  
12132 GGCACCGCGCTCGACTCGCCCTCTTGCAGCTGGCTCGCGAGGAC 12181  
1405 ArgSerLeuLysGlyLysLeuMetAsnSerLysLysLeuSerSerThrAs 1421  
||| : : : : : |||  
12482 GCTCGCTCGTGGCGCTTCGCTCGTCTCACCGAGCGCGCTCGCCAC 12231  
1421 pCysPheLysThrGluAlaPheThrSerProGluAlaLeuGlnProGlyG 1438  
||||| : : : : : ||| : : : : :  
12232 ACACCGAAGACAGCTCAAGAGACTCG.....CTCAGCGCGCGCTCG 12275  
1438 lyThrAlaLeuAlaProLysLysArgSerArgLysGlyArgAlaGlyAla 1454  
: : : : : ||| : : : : :  
12276 GGGGCTCGCGCTCGCGCAAGTGAACACGACACCTCCGCTCTTCC 12325  
1455 HisGlyLeuSerLysGlyProLeuGluLysArgProTyrLeuGlyProAl 1471  
: : : : : ||| : : : : :  
12326 TGTGCAATCG.....ACCTCAGGAGCTCCACAGGCGCTG 12366  
1471 aLeuLeuLeuThrProArgAspArgAlaSerGlyThrGlnGlyAlaSerG 1488  
: : : : : ||| : : : : :  
12367 CTAGCGGCTCGACACGAGAGACGCGCTCGCCCTCGCAACGGGA 12416  
1488 luAspAsnSerGlyGly.....LysProLysMetGluG 1502  
: : : : : ||| : : : : :  
12417 ACCCTCATCCGAGTTTGGCGAACCAACGCTCGAGCGCGCTCATCC 12466  
1494 .....GlyGlyLys.....LysProLysMetGluG 1502  
|||||..... |||  
12467 GCGCCCAAGCACCGTGGCGCTCGATTCGACCAAGGACCTTC 12516  
1502 uLeu...GlyProAlaSerGlnProGluGlyArgProCysGlnProG 1518  
: : : : : ||| : : : : :  
12517 GACGGCTCGCGCTCGTCCGACGCCCG.....AGGCCACGCGCCCT 12560  
1518 lnThrArgAlaGlnLysGlnProGlyHisThrAsnTyrSerSerTyrSer 1534  
: : : : : ||| : : : : :  
12561 GGCACAGCGCCAGTCCGATCGCGTGCACGCGGAGGCTCAACTCC 12610  
1535 LysArgLys.....ArgLeuThrArgGlyArgAl 1544  
: : : : : ||| : : : : :  
12611 GCGATGTCGTCACACCTTGGCATGTATCGGGCGACGCGCGCCGCGC... 12658  
1544 aLysAsnThrThrSerSerProCysLysGlyArgAlaLysArgArgG 1561  
: : : : : ||| : : : : :  
12659 .....TCGAGGCGAAGCGCGGATCTGTTACTGAAT 12692

```

1561 lnglnlvalleuproleuaspPro..... 1569
12693 CGGTCCAGGTGTCTCCGATACCGTACGACCGGCTGAAGGGGCT 12742
1570 ...AlaGluProGluIleArgLeuSerTyrIleSerSerCysArgLeu 1585
12743 TCGGGCAGACCTTTGGTC..... 12760
1585 uArgSerAspSerArgThrProAlaPheSerProPheValArgValGlu 1602
12761 .CCACGGCCATCGCCGACCGCCGATGATCT..... 12790
1602 ysArgSpAlaPheThrThrIleCysThrValValAsnSerProGlyasp 1618
12791 .....GCCCCATCCCCCAGCCCTGGTCT 12814
1619 AlaProGlyProHisArgGlySerSer..... 1628
12815 TCGCCCAAGCCGCAAGGTCCCATCATCTATCTACCGCTACTATGGA 12864
1629 .....SerAlaSerSerSer... 1633
12865 CTGCTGATCTCGGAGTCTGAACCCATCAAGTGTCTCATCCAGC 12914
1634 .....SerSerSerSerPheSerIleuAspAlaIleGly 1645
12915 GGGCGCGCGCGCGCTCGGACGCGCGCTTCACTCGACCGCAGCTCG 12964
1646 AlaSerIleuAlaThrIleuPro.....GlyGlySerIleuGlnPro 1659
12965 GCGCCGAGGTCTTGGCCACCGCAGTCCAGGAGTGAAGTGAAGCTCTCGC 13014
1659 cArgProSerIleuProLeuSerSerThrMetHisLeuGly...ProVal 1675
13015 GCGCTCG.....GCTTCGACATCGCAGCTCGCTCTCTCGC 13052
1675 alserIysAla...LeuSerThrSerCys 1683
13053 TGACCTGGGCTTCGACGACGACCTCTCTCG 13081

```

seq.name: /cgn2\_6/ptodata/2/lna/5A\_COMB.seq:US-08-457-342-6

seq\_documentation\_block:

Sequence 6, Application US/08457342

Patent No. 5662898

GENERAL INFORMATION:

APPLICANT: Schupp, Thomas  
 APPLICANT: Ligon, James M.  
 APPLICANT: Beck, James Joseph  
 APPLICANT: Hill, Dwight Steven  
 APPLICANT: Ryals, John Andrew  
 APPLICANT: Gaffney, Thomas Deane  
 APPLICANT: Lam, Stephen Ting  
 APPLICANT: Hammer, Phillip E.  
 APPLICANT: Uknes, Scott Joseph  
 TITLE OF INVENTION: Genes for the synthesis of  
 TITLE OF INVENTION: antipathogenic substances  
 NUMBER OF SEQUENCES: 22  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Ciba-Geigy Corporation  
 STREET: 7 Skyline Drive  
 CITY: Hawthorne  
 STATE: NY  
 COUNTRY: USA  
 ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/457,342

FILING DATE: 01-JUN-1995

```

CLASSIFICATION: 424
ERROR APPLICATION DATA:
APPLICATION NUMBER: US 08/457,205
FILING DATE: 01-JUN-1995
APPLICATION NUMBER: 08/258,261
FILING DATE: 08-Jun-1994
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-457-342-6

```

```

alignment_scores:
Quality: 493.00 Length: 2095
Ratio: 0.593 Gaps: 102
Percent Similarity: 39.666 Percent Identity: 22.291

```

alignment\_block:

US-09-508-821B-6 x US-08-457-342-6 ..

Align seg 1/1 to: US-08-457-342-6 from: 1 to: 28958

```

62 gIyAlaGlyThrProSerGlyThrAlaAlaValAla.....Al 75
7609 GCGGCGGAGACGACGACGCGCTACGCGCGCGGCGGCTTCTCGACG 7658
7659 CTTGCCGAGCATCGCGGAGCGGCTGATGAGACGACGCGGCTGCGCT 7708
90 ..GlyLeuGlnGly..ArgProAlaPheProGlyTyrGlyValGlnAsp 105
7709 GGGGCGCGTGGGCGCGCGCGGCGGCTACGCGACGACGCGCGGCGCAC 7758
105 rSer.....ProTyrProGlyArgTyrAlaGlyGln 116
7759 .CICCAACAGCGCGGTCTGCGGGAGTGGCCCGCTGCGCTGCGGCGG 7808
116 lserIleuGlnAlaThrProGlyAlaProGlnProProProGlnPro... 131
7809 GCTCCGCGTGGCTCTGACACGACGAGACGACGCTGACCGCGAGCA 7858
132 .....GlnProAla...ProAlaG 137
7859 TCGATGAGCGCGCGCTTGGCGCTTCGACGCGCTCGCGCGCGCGCC.. 7906
137 yValAlaIysTyrAspGlnAsnLeuMetIysIysThrAlaValProPro 154
7907 .....GCGTCTCGCGCGGATTTGCCGGA 7928
354 eArgGlnTyrAlaGlnGlnGlnGlnValProPheArgThrHisSer 170
7929 GCGCGACGCGCGCTCTGACACGACGAGAGCG..... 7960
171 LeuHisValGlnGlnProProProProGlnGlnProLeuAlaTyrPro 187
7961 .....GCTCTCCGACGATGCGCGCGCG..... 7984
187 sLeuGlnArgGlnIysLeuGlnAsnAspIleAlaSerProLeuProPhe 204

```

7984 ..... 7984  
204 roGlnGlyThrHisPheProGlnHisSerGlnSerPheProThrSerSer 220  
|||||  
7985 .....CGAAGCTCTCTCG 7996  
221 ThrTyrSerSerSerValGlnGlyGlyGlnGlyAlaHisSerTyrLy 237  
|||||  
7997 ACAAGCTCCGAGCGCTCGAGAGGAGGAGGAGCTCTCTCTCTCTCTCG 8046  
237 sSerCys.....ThrAlaProThrAlaGlnProH 247  
: |||  
8047 CTGTGGCCAGCAGACGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8096  
247 isAspArgProLeuThrAlaSerSerSerLeuAlaProGlyGlnArgVal 263  
: |||  
8097 TGTCAGCCCGCAGCAAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8125  
264 GlnAsnLeuHisAlaTyrGlnSerGlyArgLeuSerTyrAspGlnGln 280  
8125 ..... 8125  
280 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnSerArgH 297  
8125 ..... 8125  
297 isHisAlaGlnGlnThrLeuHisTyrGlnAsnLeuAlaLysTyrGlnHis 313  
: |||  
8126 .....TCGGTCTCGATTCGCTCA 8143  
314 TyrGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 328  
8144 TGGCGCTCGAGCTGCGCGGCTGTCGACAGGCGCCGCGCAGCTCAAGCTC 8193  
328 lArgThrProGlnGln.....TyrTyrGlnThrPheSerP 340  
: |||  
8194 CGGCGCCT 8243  
340 roSerSerHisSerProAlaArgSerValGlyArgSerProSerTyr 356  
: |||  
8244 CTTCGGGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8290  
357 SerSerThrProSerProLeuMetProAsnLeuGlnAsnPhe...ProTy 372  
8291 .....AGCCCGAGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 8328  
372 rSerGlnGlnProLeuSerThrGlyAlaPheProAlaGlyIleThrAspH 389  
: |||  
8329 AGCGAGGAGCCATCGCATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8378  
389 isSerHisPheMetProLeuLeuAsnProSer...ProThrAspAlaThr 404  
: |||  
8379 CGTCGGGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8428  
405 SerSerValAspThrGlnAlaGlyAsnCysLysProLeuGlnLys...As 420  
8429 GGGTGT.....AGCCCGATTCGACAGGCGCGG 8454  
420 pLysLeuProGlnAsnLeuLeuSerAspLeuSerLeuGlnSerLeuThrA 437  
8455 TGGGATCGCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8504  
437 lAlaLeuThrLeuGlnValGlnAsnIleSerAsnThrValGlnGlnLeu 453  
8505 CTACGTCTGGGATCGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8554  
454 LeuSerTyrAlaAlaValProGlnLysGlyValLysAsnLeuValSer 470  
8555 CTTCTTGGGATCGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8589  
470 rArgThrProGlnGlnHisLysSerGlnHisCysSerProGluGlySerG 487  
8590 CTGAGCCGCC.....AGCAGCGCTCTCTCTCTCTCTCTCTCTCTCT 8618

487. lyTyrSerAlaGluProAlaGlyThrProLeuSerGluProPro..... 501  
8619 ATCA 8668  
502 .....SerSerThrProGlnSerThrHi 509  
8669 AGGATTCCTCCACCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8718  
509 salaGlu.....ProGlnGluAlaA 516  
8719 TGGCAGAGGCGGAGCAGCAGAGATTCGAGCTTATGCTCTCTCTCTCA 8767  
516 spTyrLeuSerGlySerGluAspProLeuGlnArgSerPheLeuTyrCys 532  
8768 .....CGCGCGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8803  
532 AsnGlnAlaArgGlySerProAlaArgValAsnSerAsnSerLysAlaLy 549  
8804 CGCT 8850  
549 sProGluSerValSerThrCysSerValThrSerProAspAspMetSerT 566  
8851 TCT 8900  
566 hTyrSerAspAspSerPheGlnSerLeuHisGlySerLeuProLeuAsp 582  
8901 GTGCACT 8950  
583 SerPheSerLysPheValAla.....GlyGluArgAspCysP 595  
8951 GCT 9000  
590 .....GlyGluArgAspCysP 595  
9001 AGAGCT 9050  
595 roArgLeuLeuLeuSerAlaLeu.....lArgGlnGluAspLeuAla 608  
9051 GTCCT 9100  
609 SerGluIleLeuGlyLeuGlnGlnAlaIleGlyGlyLysAlaLysAl 625  
9101 TCT 9150  
625 a.....TrpAlaGluAlaProS 631  
9151 GGTATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9200  
631 erLeuValLysAspSerSerLysProProPheSerLeu..... 643  
9201 GCT 9250  
644 .....GluAsnHisSerAlaCysLeuAspSerVa 653  
9251 ATGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9289  
653 lAlaLysSerAlaThrProArgProGlyGluProGluAlaLeuProAspS 670  
9250 .....AGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 9325  
670 erLeuGlnLeuAspLysGlyValAsnAlaLysAspPheSerProGlyLeu 686  
9326 .....CTGAAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 9349  
687 PheGluAspProSerValAlaPheAlaThrProAspProLysLysThr 703  
9350 CGCTCAAGACCAACATCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 9399  
703 rGlyProLeuSerPheGlyThrLysProThrLeuGlyValProAlaProA 720  
9400 GTGCGCAAGATCG.....TCGCTCTCTCTCTCTCTCTCTCTCTCTCT 9443



```

720 sProthrThrAlaAlaPheaspCysPhePro.....AspThr 732
|||||
9444 CACCTCCACAGGCGCGCCGCAAT...CCCTGATGATGGATACA 9489
|||||
733 ThrAlaIAserSerAlaAspSerAlaAsnProPheAlaIAserProLue 749
|||||
9490 CACGCAATCGACGCTGATGATACC.....CCGAGCTCTGGGCGCGCA 9533
|||||
749 uAsnLeuGlyAspAlaCysProArgTrpGlyLeuHisProGlyLue 766
|||||
9534 GAA.....GATAGCACTCCCGC..... 9552
|||||
766 hTrysGlyLeuGlnGlnGlyLysAlaSerAspGlyIleSerLysGly 782
|||||
9553 .....GCGCGCGCTCTCGCCTCGGACTCGCGGACC 9588
|||||
783 AspThrHis.....GlnAlaSerAlaCysLeuGlyPheGln 795
|||||
9589 AACGCCACGTCATCTCGAGAGAGCTCCGCGCGCTG.....TC 9629
|||||
795 uGlnAspProProGlyGlyLysValAlaSerLeuProGlyAspPheLys 812
|||||
9630 GCGGAGCGCGCCACCTCACAGAGCGGTGCGAGCGCTCCCGCGCGCT 9679
|||||
812 hngLglnValGlyGlyValLysGlnGlnGlyGlyLeuLeuGlnCys 828
|||||
9680 GTGCGGTGCTCTGTGCGAGAGAGCGCGCTCCCGCGCGCGCT 9729
|||||
829 ProGln.....ValAlaLysAlaAspArgTrpLeuGln 840
|||||
9730 AACGGCTCCGCGCACCACTCTCCGCGAGAGAGCTCCGCGCTTATGCA 9779
|||||
840 pSerArgHisCysCysSerThr.....AlaAspPhe GlyAspLeuPro 854
|||||
9780 TGTGCGCTATTCGAGCGACACCGCGCGCTGCGAGCGCGCGCT 9829
|||||
855 LeuLeuProProThrSerArgLysGlnAspLeuGlnGlnGlnGly 871
|||||
9830 CTCTCTGCGCGCGCGCGCGAG..... 9853
|||||
871 rSerSerLeuGlnGlnLeuGlnGlySerProGlnGlnArgProGlyMet 888
|||||
9854 .AGCTCTCTCCGCGCTCGACTCGCTCCGCGAGAGCAAGCG..... 9895
|||||
888 hAspProLeuSerProLysAlaProLeuIleCysThrLysGlnGlnVal 904
|||||
9896 .....CCCCGAGACGCTCTCGCGCGAGCGAGCGAGCGAGCG 9931
|||||
905 GlnGlnValLeuAspSer.....LysAla.....GlyTr 914
|||||
9932 GCAAGCTCTCTCTCTCTCTCCGCGCAAGCTCGAGTGGAGAGGATG 9981
|||||
914 pGlySerProCysHisLeuSerGlyLysValIleLeuLeuGlyProT 931
|||||
9982 GCGCTCTCCCTCGACTCTCGCGCGCTCCGCGCTCGACTGAGAGC 10031
|||||
931 hValGlyThrGlnSerLysValGlnSerTrpPheGlnSerSerLeu 947
|||||
10032 ATGGAGCGCGCGCTCGCTCTCACTG..... 10060
|||||
948 HisMetLysProGlyGlnGlnLysProAsp.....GlyLysAlaLys 962
|||||
10061 .....AGTGAAGCTCTCGCGCTCGCGCGCGAGAGGCGCGC 10101
|||||
962 oGlyAspSerThrThrSerAspAla..... 970
|||||
10102 CCTCTCTCGAGCGCGCTCGAGCTCGAGCGCGCGCTCTGCGCTCAT 10151
|||||
971 .....SerLeuAlaGlnLysProAsnLysProAlaVal 981
|||||
10153 GGTCTCTCTGCGCGCGCTCTGCGCTCTGCGCGCTCGAGCGCGCGC 10201
|||||
982 ProGlnAlaProIleAlaLysLysGlnProValProArgGlyLysSer 998

```

```

10202 TCCTGCGCGCACAGCGAGCGAGATCGCGCGCGCTCTGTCGAGCGCGCT 10251
|||||
998 uArgSerArgArgValHisArgGlyLeuProGlnAlaGlnAspSerPro 1015
|||||
10252 CTCTGCTCGAGAGAGCGCGCGCGCTCGCG.....CCCT 10286
|||||
1015 ys.....ArgAlaProValLeuProLysAspLeuLeuProGln 1028
|||||
10287 GCGCAGAGAAAGCTCACACAGCTCGCGCGCGAGCGCGCATGCGCGC 10336
|||||
1029 SerCysThrGlyProProGlnGln.....MetGlnGlyAla.. 1041
|||||
10337 TCAGCTCGCGCGCTCTCGAGCTCGAGCTCACTCACTCTCTCTCGCGCGC 10386
|||||
1042 .....GlyAlaProGlyArgGlyAlaSerGln 1051
|||||
10387 AGGCTCTCACCGCGCGCGCTCGAGCGCGCGCGCTCACTCACTCTCTCGCG 10436
|||||
1051 LysLeuProArgMetCysThrArgSerLeuThrThrProAlaLeuSerGlnProArg 1067
|||||
10437 CAGAGCGCGCGCGCTCGAGCGCGCTCGAGCTCACTCACTCACTCACTCAAG 10486
|||||
1068 .....ThrProGlyProProGlyLeuThrThrThrProAlaPro..... 1080
|||||
10487 TGTTCGCGCGCAAGTCCGCGCTGACTACGCTCTCCCTCCGCGCGCATG 10536
|||||
1081 .....ProAspLysLeuGly 1086
|||||
10537 GAGCGCGTCCAGAGAGAGCTCGCGCGAGCTGAGTCAAGTCACTCTCTCTCG 10586
|||||
1086 LysLysGlnArgAlaAlaPheLysSerGlyLysArgValGlyLysProSer 1102
|||||
10587 GACGTGCGAGCTCCCTCTTAT.....GAGCGCTCA 10618
|||||
1103 ProLysAlaIAser...SerProSerAsnProAla..... 1113
|||||
10619 CCGGACACAGGCTCGAGCGCTCGAGCTCGAGCGCGCTCACTGATGCA 10668
|||||
1114 .....AlaLeuProValAlaSerAspSerProMetGlySerLysT 1128
|||||
10669 AACCTCGCGCAACCGCTCTCTCTCTGAGCGAGCGAGCGCGCTCTGCA 10718
|||||
1128 hTrysGlnThrAspSerProSer..... 1135
|||||
10719 CGATGGCATCGCTCTCTCTCTGAGCTGACACCGCGCTCTCTCTCTCTCT 10768
|||||
1136 ThrProGly LysAspGlnArgSerMetLeuArgSerArgThrLysT 1152
|||||
10769 TCGCGTCCGCGAGAGCTCGAGCGCTCACTCACTGATCCGCTCTCTCTCT 10818
|||||
1152 hTrGlnGlnIlePheHisSerLysArgArgArgProSerGlnGlyArgLeu 1168
|||||
10839 GG.....CTCACTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 10849
|||||
1169 ProAsnCys.....ArgAl 1173
|||||
10850 CCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10899
|||||
1173 ThrTrysLysLeuLeuAspAsnSerHisLeuProAlaThrPheLysVal 1190
|||||
10900 GACTGGAAGACTTCTCTCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10949
|||||
1190 rSerSerProGlnLysGln.....GlyArgValSerAlaVal 1203
|||||
10950 CACCTAACCTCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 10999
|||||
1204 ArgVal..... 1205
|||||
11000 AACGCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 11049
|||||
1206 .....ProLysProGlyAla..... 1210

```

11050 CTGCTCGGCGCGCGCTTCGCTTCGCGCAGCGGCTGCTTCCTTTTAC 11099  
1211 .....GlySerLysLeuSerAspArg 1217  
11100 AGGCGGCTCTCCCTCCAGAGACCCGCTGGCTCGAGGCCATCGCTCT 11149  
1218 ProLeuHisAlaLeuLysArgLysSerAlaPhe..... 1228  
11150 TCGGCACACCATCTACCGGCGACCGCTTCGAGCTCGCGCTGCAC 11199  
1229 .....MetAlaProValProThrLys.....LysArg 1238  
11200 GTGCCCCAGCGCTCGCTCCAGACCTCGAGAGCTCAGCTCGAGGC 11249  
1238 snLeuValLeuArgHisLysSerSerSerSerSerSerSerSerAlaMet 1254  
11250 CCGCTCGCTCTCCCATCGCAGGACACCGCTCTCTCCAGATCTCGCTG 11299  
1255 Gly..... 1255  
11300 GGCGCTGAGCAGCAGGAGGAGGCGCTCTCTTCATAGCGGACAA 11349  
1256 .....GluMetG 1258  
11350 GAGGAGCGCTTCAGGATGGCCCTGGACTGGCCAGCGCGGCTCTCT 11399  
1258 LysArgArgGlyLeuArgValProProProSerSerArgGlyCysLeu 1274  
11400 CTGCGCGGAGCCCATCCCTCTCGCGGATCTCCAGAGTGC..... 11443  
1275 LeuProArg.....LysProSe 1280  
11444 CTCCCTCGAGTCCATCCCGGTGGACTCGAGGCTCTACGCAACCTC 11493  
1280 rPro.....ProArgAlaMetAlaSerLeuProG 1290  
11494 GCCAACCTCGGCTTCGCTACGCGCCGAGTTCAGGCGCTCGGCTCGT 11543  
1290 lnSerSer.....His 1293  
11544 CTACAGCGCGGACAGCTCTTTCGAGGACCAAGCTCCCGGAGCGG 11593  
1294 ProArgArg...ProProPheLeuProGlnAlaLeuSerAlaAlaPhe 1309  
11594 CCGAAAGAGTCCGCGCGCTTTGGCC...TCCACCTCGCTCGCTCGAC 11640  
1309 eGlnGlyAlaMetLysThrLysValLeuProProArgLysGlyArgGlyL 1326  
11641 AGCGCCCTCATGCACTGGCTTTGAGGACGACGAGGAGGAGCTGCG 11690  
1326 eu..... 1326  
11691 TCTGCCCTTCGTGGAGCGGAGTCTCGCTCGCTCGCTCGCTGCCACCA 11740  
1327 .....LysLeuGluAlaIleValGlnLysIleThrSerProSer..... 1339  
11741 CCTTGGCGCTGCTTCACCGCTCCCAAGGCTGAATCTCTCGCTCGATC 11790  
1340 .....LeuLysLysPheA 1344  
11791 GTCCCTGGCGAGCGCGCAGGTGACCTTCCTTCGCTCGCTGCAAGCGCTGC 11840  
1344 laCys..... 1345  
11841 CATCGGAGGAGTCCGCGCGAGCTCCGACCCCGGAGCTTCCCAACC 11890  
1346 .....LysAlaPro..... 1348  
11891 ATGATGGCTCTTCGCTGCTGAGTGGAGGAGCTCCAAAGCCCACTCA 11940  
1348 ..... 1348  
11941 CCGCTCGCGCGCGGCGCTCTTCTCGGACAGGCGCGGCGGAGTCT 11990

1349 .....GlyAlaSerProGlyAsnProLeuSer...ProSerLeu 1360  
1199A CGGCTCGACGCCCGCTCGCTAGCGGACCTCGCTGCCCTCCGAA 12040  
1361 SerAspLysAspArgGlyLeuLys...GlyAlaGlyLysProValGI 1376  
12041 GCGCTCGACGAGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 12085  
1376 yValGlnGlyLeuValAsnValGlyThrGlyGlnLysLeuProThr. 1392  
12086 ...CCTTCATCGATCGACCGGAGGAGCTCTCGCGGCGGCGGCGG 12131  
1393 .....SerGlyAlaAspProLeuCysArg.....AsnProThrAsn 1404  
12132 GGCACCGGCTCGCACTCGCTCTTCAAGCTGCTCGCGGAGCAAC 12181  
1405 ArgSerLeuLysGlyLysLeuMetAsnSerLysLysLeuSerThrAs 1421  
12192 GCCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 12231  
1421 pCysPheLysThrGluAlaPheThrSerProGluAlaLeuGlnProGlyG 1438  
12232 CACACGAGAGAGCTCAAGGAGCTCG.....CTCAGCGCGGCTCG 12275  
1438 LysThrAlaLeuAlaProLysLysArgSerArgLysGlyArgAlaGlyAla 1454  
12276 GGGCTCGCGGCTCGCGGCTCGCGGCTCGCGGCTCGCGGCTCGCT 12325  
1455 HisGlyLeuSerLysGlyProLeuGluLysArgProLysLeuGlyProAl 1471  
12326 TCGTGCAGATCG.....ACCTCAGGAGGCTCCCGAGGCGCTG 12366  
1471 aLeuLeuLeuThrProArgAspArgAlaSerGlyThrGlnGlyAlaSerG 1488  
12367 GTAGCGGCTCGACACAGAGAGAGGCTCGCTCGCTCGCTCGCTCG 12416  
1488 LysAspAsnSerGlyGly..... 1493  
12417 ACCCTCATCGGAGTGGCGACACGCTCGCGGAGCGGCTCATCC 12466  
1494 .....GlyGlyLys.....LysProLysMetGluGI 1502  
12467 CGCGCAGACACCGCTCGGCTCGCTATTCGACCAAGGCGCTTC 12516  
1502 uLeu...GlyProAlaSerGlnProGluGlyArgProCysGlnProG 1518  
12517 GAGCGCTCGCTCGCTCGGCGCGG.....AGCGCGGCGGCGCT 12560  
1518 lnThrArgAlaGlnLysGlnProGlyHisThrAsnTyrSerSerTyrSer 1534  
12561 CGCACAGGCGCAAGTCCGCTCGCTCGCGGAGGCTCAACITCC 12610  
1535 LysArgLys.....ArgLeuThrArgGlyArgAl 1544  
12611 GCGATCTCGTCCAGACCTTCGCTGCTCGCGGCGGCGGCGGCGG... 12658  
1544 aLysAsnThrThrSerSerProCysLysGlyArgAlaLysArgArgG 1561  
12659 .....TCGAGGCGGAGGCGGCGGCTGCTTACTGAGT 12692  
1561 lnGlnGlnValLeuProLeuAspPro..... 1569  
12693 CGGTCCAGGTCTCCGATACAGCTAGCGGCGGCTGATGGGTCT 12742  
1570 ...AlaGluProGluIleArgLeuLysTyrIleSerSerCysLysArgLe 1585  
12743 TCGGCGGCGCTTTGGTC..... 12760  
1585 uArgSerAspSerArgThrProAlaPheSerProPheValArgValGluL 1602  
12761 CCACGCGCATCGCGGAGCGGCTGATCT..... 12790

```

TELEFAX : 919-541-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ANTI-SENSE: NO
US-08-457-646A-6

alignment_scores:
Quality: 493.00      Length: 2095
Ratio: 0.593        Gaps: 102
Percent Similarity: 39.66   Percent Identity: 22.291

Alignment_block:
US-09-508-821B-6 x US-08-457-646A-6 ..

Align seg 1/1 to: US-08-457-646A-6 from: 1 to: 28958

62 GLVALGLYTHPROSeGlyTHRAlAAAlAValala.....Al 75
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7609 GGCGGGGAGACCAAGAGGGGGCTACGGGGCCGAACAACGCCTCTCGACGC
75 aaSLySTYRHIsarGlySerLySaLaLeuProThrgInglN..... 89
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7659 CCTGGCGAGCATTGGGGCGACGCCGTGGATTAAGAGAGCATCGTGCGCT
90 .glyleugInly ArgProAlaIhePheoglyTyrlYValAlnaspse 105
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7709 GGGGGCGGCGGGGGCGCGCGCGCATGGCACCGACGAGGGGAGGCCAAC
105 rser.....ProtyrProdLYArGYALaglylgln 116
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7759 CTCACACAGCGGGCTCTCGCGGAGTGGCCCCCTTCCTCGCTGGCGAGC
116 InserleugInalAtPrIpLyAlAProGlnProProProGlnPro... 131
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7809 GCTGGCGCTGGCTCTGGAGACACGACAGACACACGCTACCGTGGCGGAGA
132 .....GlnProLeu..ProkIacI 137
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7859 TCGACGGGGCGGCGCTTTGCGGCTTCGTGACGGCGCGCGCCGCCGCC.. 7906
137 YVAlAlaLyTYraSPgInuEndeuleWtLysrThAlaValProPso 154
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7907 .....CGCTCCGGCGCATTTGCCCA 7928
154 eArgInLyTRAlaglIngInLyAlaGlInVaIProphearythrHisser 170
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7929 GGCGACGCGGCTCTCGAGACAGACGAGAGG.... 7960
121 LeuHIsValGlnLPProProProProGlnGlnProLeuAlatyrProLy 187
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7961 .....CGTCCCGGAGCANTGGCCGGCCC..... 7984
187 sLeugInArGynLySLeugInAsnSPIleaLaserProLeuProPhaf 204
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7984 roGInglyThrHisPheProGlnHisserGlnSerPheProTrnSerSer 220
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7985 .....CGACCTCTCG 7996
221 ThrTySerSerSerValGInglyLyGInglyAlaHisSerTyry 237
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7997 ACAAGCCCGAGACCGCGCGAGAGAGGAGACTTCGCTGCTGCGCTGCG 8046
237 sSerCyS.....ThralProThAlaGlnProH 247
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

8047 CTGGTCGCCAGGAGCGCCCTCTCTCTCGCCACGAGGCGCTCCCA 8096  
247 isAspArgProLeuThrAlaSerSerLeuAlaProGlyGlnArgVal 263  
8097 TGTCCAGCCGACGAGGCTCTCTCGATC..... 8125  
264 GlnAsnLeuHisAlaTyrGlnSerGlyArgLeuSerTyrAspGlnGln 280  
8125 ..... 8125  
280 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 297  
8125 ..... 8125  
297 isHisAlaGlnGlnThrLeuHisTyrGlnAsnLeuAlaLysTyrGlnHis 313  
8126 .....TCGGTCTCGATTCTGCTCA 8143  
314 TyrGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 328  
8144 TGGCCCTCGAGCTCGCCGCGCTTGCACAGAGCCACCGCATCAGCTC 8193  
328 larGThrProGlnGln.....TyrTyrGlnThrPheSerP 340  
8194 CCAGCCACCTCGCTCTCGACCATCCCTCTCTCATCGAGTCGCTCTT 8243  
340 roSerSerSerHisProAlaArgSerValGlyArgSerProSerTyr 356  
8244 CTGCGCGACTCGTCGCCGCCACGCCCTCGGCACAGAGCTCTCCGFCG... 8290  
357 SerSerThrProSerProLeuMetProAsnLeuGluAsnPhe...ProTyr 372  
8291 .....AGCCGAGCGCGCGCGCTCCGCGGCTTCGCGCGCG 8328  
372 rSerGlnGlnProLeuSerThrGlyAlaPheProAlaGlyIleThrAspH 389  
8329 AGCGAGAGCCCATCGCATCGCGCATGCGCTCGCGCTCGCGCGCG 8428  
389 isSerHisPheMetProLeuAsnProSer...ProThrAspAlaThr 404  
8379 CGTCGGGAGTGCAGCGCTTTTGGAGTGTCTCGGCCAGGACCGGAG 8428  
405 SerSerValAspThrGlnAlaGlyAsnCysLysProLeuGlnLys...As 420  
8429 GCGTCG.....AGCCCATTCACAAAGCCCGA 8454  
420 pLysLeuProGluAsnLeuLeuSerAspLeuSerLeuGlnSerLeuThrA 437  
8455 TGGATCGCGCTCGCGCTCTACGACCGCGCGCGCGCGCGCGCGCG 8504  
437 laLeuThrLeuGlnValGluAsnIleSerAsnThrValGlnGlnLeuLeu 453  
8505 CTAGCTCCGCGATGCGCCCATGTGCGACGATGCGAGCTCTTCGACCTG 8554  
454 LeuSerIysAlaAlaValProGlnLysLysGlyValIysAsnLeuValSe 470  
8555 CTTCTCTTGGCATCAGCCCGCGGAGGCAAC.....AC 8589  
470 rArgThrProGlnGlnHisLysSerGlnHisCysSerProGluGlySerG 487  
8590 CTCAGCCCC.....AGCACCGCTCTGCTC.....TCGA 8618  
487 lyTyrSerAlaGluProAlaGlyThrProLeuSerGluProPro..... 501  
8619 ATCTGCTGCGAGGCTCTGAGAGAGCGCGCATGTCCTCCCGCCACCTCA 8668  
502 .....SerSerThrProGlnSerThrHI 509  
8669 AGGATTCCTCCACCGCGTCTCTGTCGCGCATCGCGCGCGCATACGCA 8718  
509 sAlaGlu.....ProGlnGluAla 516  
8719 TTGCGAGGCGAGCGAGGATTCGAGCGCTTATGCTCCCTCCAGGCA. 8767

516 spTyrLeuSerGlySerGluAspProLeuGluArgSerPheLeuTyrCys 532  
8748 .....CCGCGGCTCTTTTGGCGGCGCTTGGCTACA 8803  
533 AsnGlnAlaArgGlySerProAlaArgValAsnSerAsnSerLysAlaLys 549  
8804 CGTTCGCGCTCGAAGGCGCG...CGCTCTCGGTCGACACCGCTCTGCTCC 8850  
549 sProGluSerValSerThrCysSerValThrSerProAspAspMetSerT 566  
8851 TCCTCGCTCTGCGCTCCACCTCGCTGCCAAGCCTCCGACAGGCGGA 8900  
566 hrLysSerAspSerPheGlnSerLeuHisGlySerLeuProLeuAsp 582  
890A GTGCAACCTCGCTCGCGCGGCTCTCGGTCTATGGCTCTCCCGGAG 8950  
583 SerPheSerLysPheValAla..... 589  
8951 GCTTCGCTCTCTTTCCGCTCTCGGCTCTGGCGCTTGGCGCCGCGCTCC 9000  
590 .....GlyGluArgAspCysP 595  
9001 AAGACCTTCTCGGCAAGCGCGAGCTACGAGCGGAGAGGCGTCTAT 9050  
595 roArgLeuLeuSerAlaLeu.....AlaGlnGluAspLeuAla 608  
9051 CGTCTTCGCTCGAGGCGCTCGGTGAGCGCTCTCGCGGAGGACACCGCG 9100  
609 SerGluIleLeuGlyLeuGlnGlnAlaIleGlyGlyLysAlaAspLysAl 625  
9101 TCCTCGCTCTGTCGCGGCGCGCATCAACACGACGCGCGCTCGAGC 9150  
625 a.....TrpAlaGluAlaProS 631  
9151 GGTATCAGCGCGCCCAAGCGCACCTCCAGCAGAGAGTCTCTCGCGCGC 9200  
631 erLeuValLysAspSerSerLysProProPheSerLeu..... 643  
9201 CTCTCAGCGCGCGCTACCGCGCGCTCGACGCTCGAGTCTCGAGTCC 9250  
644 .....GluAsnHisSerAlaCysLeuAspSerVa 653  
9251 ATGGCAGCGGACCTCTCTGGGAGACCGCATCGAGTGC..... 9289  
653 laLysSerAlaThrProArgProGlyGluProGluAlaLeuProAspS 670  
9290 .....AAGCCCTGCGCG...CCGTCTACGCGCGCGCGAGACCGCG... 9325  
670 erLeuGlnLeuAspLysGlyAsnAlaLysAspPheSerProGlyLeu 686  
9326 .....CTGAAAGCTCTCTCTCTCGCGG 9349  
687 PheGluAspProSerValAlaPheAlaThrProAspProLysLysThrTh 703  
9350 CGCTCAGACCAACATCGGCATCTCGAGGCGCGCTCGGCTCTCGCGCG 9399  
703 rGlyProLeuSerPheGlyThrLysProThrLeuGlyValProAlaProA 720  
9409 GTGCCAGAGATG.....TCGCTCTCTCTCGGCATGAGCGCTCTCGCCCC 9443  
720 spProThrThrAlaAlaPheAspCysPhePro.....AspThr 732  
9444 QACCTCCACAGCGCGCGCGCAAT...CCCTTGATGATGAGGATACA 9489  
733 ThrAlaSerSerAlaAspSerAlaAsnProPheAlaThrProGluG 749  
9490 CTCGCCATCGAGTCTGTGATACC.....CCGAGGTCTTGGCGCGCGCA 9533  
749 uAsnLeuGlyAspAlaCysProArgTrpGlyLeuHisProGlyGluLeu 766  
9534 CGAA.....GATAGAGTCCCGCG... 9552

766 hrlselyleuclnglylvalaseraapgllyserlysgly 782  
9553 .....  
783 AspThrHis.....GluAlaSerAlaCysLeuGlyPheGlnI 795  
9589 AACGCCACAGTCATCTCGAGAGAGGCTCCGCCCTG.....TC 9629  
795 uGlnuapProGlyGlyLysValAlaSerLeuProGlyAspPheLysG 812  
9630 GGGCGAGCCGCCACCTCACAGAGCGCTCCGACCGCCCGGCGCT 9679  
812 lngluGlnValGlyLysValGlyLysValGlyLysValGlyLys 828  
9680 GTGCCGTGCTCTGTCGCGAGAGAGCGCTCCGCCCGCCACGCG 9729  
829 ProGln.....ValAlaLysAlaAspArgTrpLeuGlnAs 840  
9730 AACGGCTCCGCGACACCTCTCTCCGACAGACCTCCGCTTATCGA 9779  
840 pSerArgHisCysCysSerThr.....AlaAspPhe.GlyAspLeuPro 854  
9780 TGTGGCTATTTCGAGGCGACACCGCGCCACTTCGAGCGCGCG 9829  
855 LeuLeuProProThrSerArgLysGlnAspLeuAlaGlnGlnLys 871  
9830 CTCTCTGCGCCCGGACCGGAG..... 9853  
871 rSerSerLeuGlyGlnLeuGlySerProGlnLysArgProGlyMetG 888  
9854 .AGCTCTCTCCGCGCTCGACTCGCTCCGCGACGACACACCGCG..... 9895  
888 lnsAspProLeuSerProLysAlaProLeuLeuGlyThrLysGlnVal 904  
9896 .....CCCCGAGACAGCTTCTGCGCGGAGCGAGCGACCGAC 9931  
905 GlnGlnValLeuAspSer.....LysAla.....GlyTrp 914  
9932 GCAAGGTGCTCTGCTCTTCTGAGCAAGGCTCGACAGTGAAGGATG 9981  
914 pGlySerProCysHisLeuSerGlyLysValLeuLeuGlyProT 931  
9982 GCCCTCTCCCTGCTCGACTCTCGCGGCTTCGCGCTCTACTGAGAAC 10031  
931 hValGlyThrGlnSerLysValGlnSerTrpPheGlnSerSerLeuSer 947  
10032 ATGGAGGCGCGCGCTCGCTCTCACTG..... 10060  
948 HisMetLysProGlyGlnGlnGlyProAsp.....GlyGlnAlaArg 962  
10061 .....AGTGAAGCTCTCGCGCTCTCGCGCGACGAGGCGCGC 10101  
962 oGlyAspSerThrHisSerAspAla..... 970  
10102 CCTCCCTCGACCGCGCTCGAGCTCGACAGCCCGCTCTTGGCGCTAT 10151  
971 .....SerLeuAlaGlnLysProAsnLysProAlaVal 981  
10152 GGTCTCCCTGCGCGCGCTCTGCGCTCGCGCTCGAGCGCGCGCGC 10201  
982 ProGlnAlaProLeuAlaLysLysGlnProValProArgGlyLysSerLe 998  
10202 TCGTCGCGCACAGCGAGCGAGATCGCGCGCTTCGTCGAGGCGCT 10251  
998 uArgSerArgArgValHisArgLysLeuProGlnAlaGlnAspSerProC 1015  
10252 CTCTCCCTCGAGAGAGCGCGCGCTCGCG.....CCT 10286  
1015 ys.....ArgAlaProValLeuProLysAspLeuLeuProGln 1028  
10287 GCGCAGAGAAAGCGCTCACACCTCGCGCGCGACGCGCGAGCGCGCG 10336  
1029 serCysThrGlyProProGlnGlyGln.....MetGlnGlyAla.. 1041

10337 TCAGACCTGCGCGCTCCGACCTTCACCTACCTCCCGGCGAGAC 10386  
1042 .....GlyAlaProGlyArgGlyAlaSerGln 1051  
10387 AGGCTCTCCACCGCGCGCTCAACAGCCCGAGGCTACCTCGATTCGG 10436  
1051 LysLeuProArgMetCysThrArgSerLeuThrAlaLeuSerGlnProArg 1067  
10437 CGAGCCCGCGCGCGCTCGAGCGCTCTCGAGCTCTCCACCGCACAG 10486  
1058 .....ThrProGlyProProGlyLeuThrThrProAlaPro..... 1080  
10487 TGTTCGCCGCAAGTCCGCGTGCAGTTCGCGCTCCGCGCAAGT 10536  
1081 .....ProAspLysLeuGly 1086  
10537 GAGCGCTCCAGACAGCTCGCGCGAGGTCTACCAATCGCTCTCG 10586  
1086 LysGlnLysArgAlaAlaPheLysSerGlyLysArgValGlyLysProSer 1102  
10587 GAGTCCGAGCTCCCTCTTAT.....CGACCTCA 10618  
1103 ProLysAlaAlaSer.....SerProSerAsnProAla..... 1113  
10619 CCGGACACAGGCTCGAGGCTCGAGCTCGAGCGCGCTGACTGTATCGA 10668  
1114 .....AlaLeuProValAlaSerAspSerSerProMetGlySerLys 1128  
10669 AACCTCGGCAACCGCTCTGCTTCGAGCGAGCGACGAGCGCTCTCGA 10718  
1128 hrlsGlnThrAspSerProSer..... 1135  
10719 CGATGGGACATCGCTCTCTCGAGAGTCAAGCCCAATCCGAGTCACGC 10768  
1136 ThrProGlyLysAspGlnArgSerMetLeuLeuArgSerArgThrLys 1152  
10769 TCGCCCTCGCGAGACTCGAGCGCTCGAGCGCTCGAGTCCGCTCTC 10818  
1152 hrlGlnLysLeuPheHisSerLysArgArgArgProSerGlnGlnLysLeu 1168  
10819 GG.....CTCATTCAGACAGAGAGAGCACTCG..... 10849  
1169 ProAsnGly.....ArgAl 1173  
10850 CCGCGCTCTCTCTCTCGAGCGAGCTCTACCGAGCGCTCTCGCTC 10899  
1173 aThrLysLysLeuLeuAspAsnSerHisLeuProAlaThrPheLysValS 1190  
10900 GACTGAGAGGACTCTTCGCGCGCTCTACGCTCCGCAAGGCTCTCCCTCC 10949  
1190 .aSerSerProGlnLysGln.....GlyArgValSerGlnArgAla 1203  
10950 CACTACCTCTTCAGAGAGAGAGGCTTCGCTCGAGTCTCCAGAGAGC 10999  
1204 ArgVal..... 1205  
11000 AACGCTTCGAGAGTCCCTCCGAGCGCTGACTCGCGCGACCAATCCCG 11049  
1206 .....ProLysProGlyAla..... 1210  
11050 CTGCTCGCGCGCGCTCTGCGCGCTTCGCGAGCGCGGCTCTCTTAC 11099  
1211 .....GlySerLysLeuSerAspArg 1217  
11100 AGGCGGCTCTCTCTCGAGAGACCGCGGCTCGAGAGCGATGCCGCTC 11149  
1218 ProLeuHisAlaLeuLysArgLysSerAlaPhe..... 1228  
11150 TCGGACACACCAATCTACCGGCGACCGGCTTCGAGCTCGCGCTCGAC 11199  
1229 .....MetAlaProValProThrLys.....LysArg 1238

11200 CTCGCCACCGCTCGGCTCGACACCGTCGAGAGCTACGCTCGAGGC 11249  
1238 snLeuValLeuArgHisGlySerSerSerSerSerSerSerSerSerSer 1254  
11250 CCTCTCGCTCTCCATCGACGACACCGCTCTCTCCATCGCTCGGTCG 11299  
1255 Gly ..... 1255  
11300 GCGCCGTCGACGACGACGACGACGACGACGACGACGACGACGACGAC 11349  
1256 ..... 1256  
11350 GAGGACGCGCTTCAGATGCGCCCTGAGCTCGCCACGCGCGCGCTCTCT 11399  
1258 ValArgArgArgGlyLeuArgValProProProProProProProProPro 1274  
11400 CTCGCGCGGACGACGACGACGACGACGACGACGACGACGACGACGAC 11443  
1275 LeuProArg ..... 1275  
11444 CTCCTCGAGTGCATCCCGGTGACCTCGAAGGCTTCAGCAACCCCTC 11493  
1280 rPro ..... 1280  
11494 GCAACCTCGGCGCTTCGCTACGCGCGCGGAGTTCAGGCGCTTCGCTCC 11543  
1290 InSerSer ..... 1293  
11544 CTCACG 11593  
1294 ProArgArg...ProProPheLeuProGlnAlaArgLeuSerAlaAlaPh 1309  
11594 CGCAAGAGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11640  
1309 eGlnGlyAlaMetLysThrLysValLeuProProArgLysGlyArgGlyL 1326  
11641 AGCGCCCTCGATCACTGCGCTTCGAGGACGACGACGAGGAGGCGCTCG 11690  
1326 eu ..... 1326  
11691 TCTGCGCTTCGCTGGAGCGGAGTCTCGCTCGCTCGCTCGCTCGCTCGC 11740  
1327 ..... LysLeuGlnAlaValGlnLysLysLysLysLysLysLysLys 1339  
11741 CCTTGGCGCTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCG 11790  
1340 ..... LeuLysLysPheA 1344  
11791 GTCTGGCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11840  
1344 laCys ..... 1345  
11841 CANGCGGACGAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11890  
1346 ..... LysAlaPro ..... 1348  
11891 ATGATCGCTCTTCGCGCTCGACTGAGCGAGCTCCAAAAGCCCACTTCA 11940  
1348 ..... 1348  
11941 CCGCTGCG 11990  
1349 ..... GlyAlaSerProGlyAsnProLeuSer...ProSerLeu 1360  
11991 CGCGCTCGAGCGCGCGCTCGCGCTGACGCGAGCTCGCTCGCTCGCTCG 12040  
1361 SerAspLysAspArgGlyLeuLys...GlyAlaGlySerProValGln 1376  
12041 GCGCGCTTCGACGAGGCGCTTCGCGCTTCGCGCTTCGCGCTTCGCGCT 12085  
1376 yValGlnGlyLeuValAlaValGlyThrGlyGlnLysLeuProThr. 1392  
12086 ..... CTTTCATGATCGACGCGGACGCGGACCTTCGCTCCGCGCGCGCA 12131

1393 ..... SerGlyAlaAspProLeuCysArg ..... AsnProThrAsn 1404  
12432 GCGCAGCGCGCTCGACTCGCTCTTCAGAGCTGGCTCGCGCGCGAGAC 12181  
1405 ArgSerLeuLysGlyLysLeuMetAsnSerLysLysLeuSerSerThrAs 1421  
12182 GCTCGCTCGCTCGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 12231  
1421 pCysPheLysThrGlnAlaPheThrSerProGlnAlaLeuGlnProGlyG 1438  
12232 CACACGAGACGACGCTCAAGACCTCG ..... CTCACGCGCGCTCGT 12275  
1438 LysThrAlaLeuAlaProLysLysArgSerArgLysGlyArgAlaGlyAla 1454  
12276 GGGCTCGCGCTCGCGCTCGCGCTCGCGCTCGCGCTCGCGCTCGCTTC 12325  
1455 HisGlyLeuSerLysGlyProLeuLysArgProProLysLeuGlyProAl 1471  
12326 TCGTGCATCG ..... ACCTCAGGAGGCTTCACGAGGCGCTG 12366  
1471 aLeuLeuThrProArgAspArgAlaSerGlyThrGlnGlyAlaSerG 1488  
12367 CTAGCGCGCTCGACACGAGGACGCTCGCTCGCTCGCTCGCTCGCTCG 12416  
1488 LuAspAsnSerGlyGly ..... 1493  
12417 ACCCTTCATCCGAGGTGGCGACACGCTCGACGCGCGCTCATCC 12466  
1494 ..... GlyGlyLys ..... LysProLysMetGlnG 1502  
12467 CCGCGCAGACCCACGCTCGCTCGCTCATTCGACCAAGGCGACCTTC 12516  
1502 uLeu...GlyProAlaSerGlnProGlnGlyArgProCysGlnProG 1518  
12517 GAGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 12560  
1518 LysThrArgAlaGlnLysGlnProGlyHisThrAsnTyrSerSerTyrSer 1534  
12561 CGCACGCGCGAGTTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 12610  
1535 LysArgLys ..... ArgLeuThrArgGlyArgAl 1544  
12611 GCGATGCTCGACACCTTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 12658  
1544 aLysAsnThrThrSerSerProCysLysGlyArgAlaLysArgArgG 1561  
12659 ..... TCGAGGCGAGGCGCGCTCGCTCGCTCGCTCGCTCGCTCGCT 12692  
1561 LysGlnLysValLeuProLeuAspPro ..... 1569  
12693 CCGTTCAGTGTCTCCGATACACCTGAGCGCGCTCGCTCGCTCGCTCG 12742  
1570 ...AlaGlnProGlnLysLeuLysLysLysLysLysLysLysLysLys 1585  
12743 TCGCGCGCGCTTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 12760  
1585 uArgSerAspSerArgThrProAlaPheSerProPheValArgValGln 1602  
12761 CCGAGCGCTCGCGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 12790  
1602 yArgAspAlaPheThrThrLysCysThrValValAsnSerProGlyAsp 1618  
12791 ..... GCGCATCCCGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 12814  
1619 AlaProLysProHisArgLysProSerSer ..... 1628  
12815 TCGCGCGCGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 12864  
1629 ..... SerAlaSerSerSer ..... 1633  
12865 CTCGCTCATCTCGGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 12914



8125 ..... 8125  
297 i s h i s a l a g l i n l u t h r l e u h i s t y r g l n a s n l e u a l a l y s t y r g l n h i s 313  
8126 ..... T G G T C T G A T T C G C T C A 8143  
314 T y r g l y c l i n g l i n g l y t y r c y s g l n ..... P r o s p a l a l a v a 328  
8144 T G G C G G C T G C G C G G C T T C G C G G C T T C A A C A G C C C A C C G C A T C A A G T C 8193  
328 l a r g t h r p r o g l u g l n ..... T y r t y r g l n t h r p h e s e r p 340  
8194 C G G C C A C C T C G C C T T C G A C C A C C T C T C T C A T G A G T C G C C T T T 8243  
340 t o s e r s e r h i s s e r p r o l a a r g s e r v a l g l y a r g s e r p r o s e r t y r 356  
8244 C T T C G C G A C T C G C T C G C C A C C C C T C G C A C A G G C T C T C G T C G ... 8290  
357 s e r s e r t h r p r o s e r p r o l e u m e t p r o a s n l e u g l u a s n p h e ... P r o T y 372  
8291 ..... A G C C G A C G C C G C G C T C C G C G C T T C G C G C G C G 8328  
372 r s e r g l n p r o l e u s e r t h r g l y a l a p h e p r o a l a c l y l l e t h r a s p h 389  
8329 A G C A C G A G C C C A T C C C A T G C G C A T G C C T C G C G C G C G 8378  
389 i s s e r h i s p h e m e t p r o l e u a s n p r o s e r ... P r o t h r a s p a l a t h r 404  
8379 C G T C G G A T G C A G C G C T T T T G G A G T T C C T G C C C A G G A C G C G A C G 8428  
405 s e r s e r v a l a s p h r g l n a l a g l y a s n c y s l y s p r o l e u g l n l y s ... A s 420  
8429 G C G T C G ..... A G C C A T T C C A A G C C C G A 8454  
420 p l y s l e u p r o g l u a s n l e u s e r a s p l e u s e r l e u g l n s e r l e u t h r a 437  
8455 T G G A T C C G C T G C C T T A C A G C C C G C C A G C C A A G A C C A A G A G 8504  
437 l a l e u t h r l e u g l n v a l g l u a s n l e s e r a s p h r v a l g l n g l n l e u l e u 453  
8505 C T A G C T C G C G A T G C G C C A T G C T C A C C A G C T G A C C T T C T G A C C C T G 8554  
454 l e u s e r l y s a l a v a l p r o g l n l y s l y s g l y v a l l y s a s n l e u v a l s e 470  
8555 C C T C T T T G G C A T C A G C C C G G G A G G C A A A C ..... A C 8589  
470 r a r t h r p r o g l u g l n h i s l y s s e r g l n h i s c y s s e r p r o g l u g l y s e r g 487  
8590 C T G A C C C C C ..... A G C A C C C C T G C T C C ..... T C G A 8618  
487 l y t y r s e r a l a g l u p r o a l a g l y t h r p r o l e u s e r g l u p r o p r o ..... 501  
8619 A T C T G C C T G C A G G C C C T C G A A G A C G C G G C A T C T C C C C C C A C C C T C A 8668  
502 ..... S e r s e r t h r p r o g l n s e r t h r H I 509  
8669 A G G A T T C C C C A C G G C G T C T T C T G C G C A T G C G C C A C C G A A T A C G C A 8718  
509 s a l a g l u ..... P r o g l n g l u a l a 516  
8719 T T G G A G A G C G G A C C G A G A T T C G A C G C T T A T G C C T C C A G G C A . 8767  
516 s p t y r l e u s e r g l y s e r g l u a s p r o l e u g l u a r g s e r p h e l e u t y r c y s 532  
8768 ..... C G G C G G G T C C T T T G C G C G G G C G T T G C C T A C A 8803  
533 A s n g l n a l a r g l y s e r p r o a l a a r g v a l a s n s e r i s n s e r l y s a l a l y 549  
8804 C G C T G C G C C T G C A A G G C C G . . . C G T C T G C G T C G A C C A C C G C T C T C C 8850  
549 s p r o g l u s e r v a l s e r t h r c y s s e r v a l t h r s e r p r o a s p a s p m e t s e r t 566  
8851 T C C T G C T C T C G C C T C C A C C T C G C T G C C A A G C C C T C G A C A G G G C G A 8900

566 h r l y s e r s e r a s p s e r p h e g l n s e r l e u h i s g l y s e r l e u p r o l e u a s p 582  
8901 T F G C A A C C T G C G C C T C G C G G G C C T C T C G C T C A T G C C T C C C C G G A G 8950  
583 S e r P h e S e r l y s P h e V a l a l a ..... G l y l u a r g A s p C y s p 595  
8951 C C T T C G T C C T C C T T T C C G C C T G C G G C T T G G C C C G A C G C G C G C T C C 9000  
590 ..... T y r t y r g l n t h r p h e s e r p 595  
9001 A A G A C C T T C T C G C C A C A C C G A C G C T A C G A C G C G G A G A G G C G T C A T 9050  
595 t o a r g l e u l e u s e r a l a l e u ..... A l a g l n g l u a s p l e u a l a 608  
9051 C G T C C T T G C C C T C G A G G C T C G G T G A C G C C T C G C C C G A G G A C A C C G C G 9100  
609 S e r g l u l e l e u g l y l e u g l n g l u a l l e g l y l e u l y s a l a s p l y s a l 625  
9101 T C T C G C C C T C T C G C G G C A C G C C A T C A C C A G A C G G C G C T C G A G C 9150  
625 a ..... T r p a l a g l u a l a p r o s 631  
9151 G T A T A C C G C C C C A C A G G C A C C T C C C A G A G A G T C T C G C G C G C G C 9200  
631 e r l e u v a l l y s a s p s e r s e r l y s p r o p h e s e r l e u ..... 643  
9201 C C T C A C A G A C G C C A T C A C C C C C G C A C G T C G A C T C G T C G A G T G C C 9250  
644 ..... G l u a s n h i s s e r a l a c y s l e u a s p s e r v a 653  
9251 A T G C A C C G G C A C C T C C T T G G A G A C C C A T C G A G G T G C ..... 9289  
653 l a l y s e r a l a t r p r o a r g p r o g l y g l u p r o g l u a l e u p r o a s p s 670  
9290 ..... A A G C C C T G S C G . . . C G G T C A C G C C A C G C A C C G . . . 9325  
670 e r l e u g l n l e u a s p l y s g l y l y s a l a l y s a s p s e r p r o g l y l e u 686  
9326 ..... C T G A A A G C C T C T C T C T C G G G C G 9349  
687 P h e g l u a s p r o s e r v a l a l a p h e a l a t h r p r o a s p p r o l y s l y s t h r 703  
9350 C G C T A A G A C A C A C T C G C C A T C T G A G G C G C C T C G G C C T C G G G C G 9399  
703 r c l y p r o l e u s e r p h e g l y t h r l y s p r o t h r l e u g l y v a l p r o a l a p r o a 720  
9400 G T C G C C A G A T G . . . . . T G C C T C C C T C C G C C A T A C G C C T C G C C C 9443  
720 s p r o t h r t h r a l a l a p h e a s p c y s p h e p r o ..... A s p t h r 732  
9444 C A C C C T C C A C A C G G C C C G C A T . . . . . C C T T A T T A T T G G A T A C A 9489  
733 T h r a l a l a s e r s e r a l a s p s e r a l a s n p r o p h e a l a t r p r o g l u g l 749  
9490 C T C C C A T C G A C T G T T A T A C . . . . . C G A G G T C T T G G C C G G C A 9533  
749 u a s n l e u g l y a s p a l a c y s p r o a r g t r p g l y l e u h i s p r o g l y g l u l e u t 766  
9534 C G A A . . . . . G A T A G A G T C C C G C . . . . . 9552  
766 h r l y s g l y l e u g l n g l y l y s a l a s e r a s p g l y l e s e r l y s g l y 782  
9553 ..... G C G C G G G C T C T C G C C T C G G A C T C T C G G C A C C 9588  
783 A s p t h r h i s ..... G l u a l a s e r a l a c y s l e u g l y p h e l n g l 795  
9589 A A C C C C A C A T C A T C T C G A G A G G T C C C G C C C T G . . . . . T C 9629  
795 u g l u a s p r o g l y g l u l y s v a l a l a s e r l e u p r o g l y a s p h e l y s g 812  
9630 G G G C A G C C G C C A C C T C A C A G A G G G G T C G G A C C G C T C C C G G G G T 9679





[illegible]





```
9830 CTCTCTGCGCCGCGACGCGAGC..... 9833
9871 rSerSerLeuGylLeuLeuGylSerProGlnAlaArgProGlyMetG 888
9854 .AGCTCTCTCCGCGCTGAGCTGCGCCGACGAGCAAGCCG..... 9855
888 lnaSerProLeuSerProLysAlaProLeuIleCysThrLysGlnGluVal 904
9896 .....CCCGAGACAGCTTCTGCGCGGAGGAGGAGCAGC 9931
905 GluGlnValLeuAspSer.....LysAla.....GlyTr 914
9932 GCAAGGTCGTCTCTCTTCTCTGCGCAGAGCTGCGAGTGGAGGAG 9981
914 pGlySerProCysHisLeuSerGylGluSerValIleLeuLeuGlyProT 931
9982 GCCCTCTCCCTGCTGAGCTCTGCGCGGTCTTCCGCGCTGAGCTCAAC 10031
931 hrValGlyThrGluSerLysValGlnSerTrpPheGluSerSerLeuSer 947
10032 ATGGAGAGCGGCGCTGCTCTCTCACTG..... 10060
948 HisMetLysProGlyGlnGluGlyProAsp.....GlyGluArgAlaPr 962
10061 .....AGTGGAGCTCTGCTGCGCTCTGCGCGAGCAGAGGCGGC 10101
962 cGlyAspSerThrThrSerAspAla..... 970
10102 CCTCTCTGACCGCGCTGACGCTGACAGCCCGCTCTTCTCTCTAT 10151
971 .....SerLeuAlaGlnLysProAlaVal 981
10152 GGTCTCTGCGCGCGCTCTGCGCTGCTGCGGCTGAGCGCGCGCG 10201
982 ProGluAlaProIleAlaLysLysGluProValProArgLysSerLe 998
10202 TCGTGGCGCCAGCCAGCGGCGAGTGGCGCGCTTCTGCGAGCGCT 10251
998 uArgSerArgArgValHisArgGlyLeuProGlnAlaGlnAspSerProc 1015
10253 CTCTCTCTGAGGAGCGCGCGCGATCGCG.....CCCT 10286
1015 ys.....ArgAlaProValLeuProLysAspLeuLeuProGlu 1028
10287 GCGGAGGAAAGCGCTCACCGCTGCGGCGCAAGCGCGATGCGCGCG 10336
1029 SerCysThrGlyProProGlnGlyGln.....MetGlnGlyAla.. 1041
10337 TCGAGCTGCGCGCTGCGACCTCGACCTGCTGCGCGCGCGCGAC 10386
1042 .....GlyAlaProGlyValArgLysSerGlu 1051
10387 AGGCTCTGACGCGCGCGCTGACAGCGCGCGCTGACCTGCTGATCCG 10436
1051 lYleuProArgMetCysThrArgSerLeuThrAlaLeuSerGluProArg 1067
10437 CGAGCCGCGCGCGCTGAGCGCGCTGCTGACAGCTCTCACCGCACAG 10486
1068 ..ThrProGlyProProGlyLeuThrThrProAlaPro..... 1080
10487 TGTTCGCCGCGACAGATCCGCTGACCTGACCTGCGCGCGCGCGAG 10536
1081 .....ProAspLysLeuGly 1086
10537 GAGCGCGCTGACAGAGCTGCGCGCGCTGACAGCTGACAGCTGCTCTG 10586
1086 lYlYsGlnArgAlaAlaPheLysSerLysArgValGlyLysProSer 1102
10587 GACGTGCGAGCTCTCTTAT.....CGACGCTCA 10618
1103 ProLysAlaLaser...SerProSerAspProla..... 1113
10619 CCGGACAGGCTGACGCGCTCCGAGCTGACGCGCGCTGACTGATCA 10668
1114 .....AlaLeuProValAlaSerAspSerProMetGlySerLysT 1128
10669 AACCTCGCGCAACCGCTCTCTCTGAGCGGACGAGCGCGCTCTCA 10718
1128 hrLysGlnThrAspSerProSer..... 1135
10719 CGATGGGATCGCTCTCTCCGCTGAGGTCAGCCCGCTGCTGACG 10768
1136 ThrProGlyLysAspGlnArgSerMetIleLeuArgSerThrLysT 1152
10769 TCGCCCTCGGAGACCTGCGAGCGCTGACCGCTGACCGCTGCTGTC 10818
1152 hrGlnGlnIlePheHisSerLysArgArgArgProSerGlnGlyArgLeu 1168
10819 GG.....CTCCATCTGACGAGAAAGAGCGCACTCG..... 10849
1169 ProAsnCys.....ArgAl 1173
10850 CCGCGCTCTCTCTCTCTGCGGAGCTCTCTACCGGAGCGCTGCTG 10899
1173 ThrLysLysLeuLeuAspAsnSerHisLeuProAlaThrPheLysVal 1190
10900 GACTGAGAGACTTCTTCTGCGCGCTGACGCTGCGCGAGCTCTCTCC 10949
1190 erSerSerProGlnLysGln.....GlyArgValSerGlnArgAla 1203
10950 CACCTACCCCTTCCAGGAGAGCGGCTTGTGCTGACAGCTCTCAAGAG 10999
1204 ArgVal..... 1205
11000 AACGCTTCCAGCTGCTGCTCCGACGAGCTGCTGCGCGAGCAATCCG 11049
1206 .....ProLysProGlyAla..... 1210
11050 CTGCTGCGCGCGCGCTGCTGCTGCGCGAGCGGCTTCTCTCTTAC 11099
1211 .....GlySerLysLeuSerAspArg 1217
11100 AGGCGGCGCTCTCTGCGAGAGACCGCTGCTGAGAGCGATGCGCT 11149
1218 ProLeuHisAlaLeuLysArgLysSerAlaPhe..... 1228
11150 TCGGACACCCATCTCTACCGGCGCGCTTCTGAGCTGCGCGCTGAC 11199
1229 .....MetAlaProValProThrLys.....LysArg 1238
11200 GTGCGCCACCGCGCTGCGCTGACACCGCTGAGAGCTGACGCTGAGG 11249
1238 snLeuValLeuArgHisGlySerSerSerSerSerAlaSerAlaMet 1254
11250 CCCTCTGCTGCTCTCCATCGAGAGACCGCTGCTGCTGAGCTGCTG 11299
1255 Gly..... 1255
11300 GCGCGCTGAGAGAGCAGAGAGAGGCGCTCTCTTCCATGACCGCAA 11349
1256 .....GluMetG 1258
11350 GAGAGCGCTTCAAGATGCGCTGAGCTGCGCGAGCGAGCGGCTCT 11399
1258 lYArgArgArgGlyLeuArgValProProProSerSerArgLysLeu 1274
11400 CTGCGCGGAGACCCATCCCTCTCCGCGAGCTGACAGAGTGC..... 11443
1275 LeuProArg.....LysProse 1280
11444 CTCTCTGAGTGGACATCCGCTGAGCTGAGAGCGCTTACAGCACTCT 11493
1280 rPro.....ProArgAlaMetAlaSerLeuProG 1290
11494 GCCACCTGCGGCTTGTGCTTACGCGCCGAGTGTCCAGGCGCTCGCTCG 11543
```

1290 lnsrSer.....His 1293  
: : : : :  
11544 CTACAAGCGGGGAGAGAGCTTTTGGCGAGCAAGCTCCCGGAGCGG 11593  
  
1294 ProArgArg...ProProPheLeuProGlnAlaArgLeuSerAlaAlaPh 1309  
: : : : :  
11594 CCGAAAGAGTGGCGCGCGGTTGGCC...TCCACCTCGGCTCGAC 11640  
  
1309 eGlnGlyAlaMetLysThrLysValLeuProProArgLysGlyArgLyl 1326  
: : : : :  
11641 AGCGCCGTGCATGCACTGGCCCTTGGAGGAGGAGGAGGAGGTCG 11690  
  
1326 eu..... 1326  
||  
11691 TCTGCCCTTCTCGGAGGAGTGTGCTGGCTCGCTCGGCTGCCACCA 11740  
  
1327 .....LysLeuGlnAlaIleValGlnLysIleThrSerProSer..... 1339  
: : : : :  
11741 CCTTGGCGTGGCTTCCACCTGCCAGGAGTGAATCCCTCGCTCGATC 11790  
  
1340 .....LeuLysLysPheA 1344  
: : : : :  
11791 GTCTGGCGGAGCGCGAGGTGACCTCTTGGCTCGGTCAAGGCTCGC 11840  
  
1344 lscys..... 1345  
|||  
11841 CATGGGAGAGCTCCGCGCGAGCTCCGACCCCGGAGCTTCCCAAC 11890  
  
1346 .....LysAlaPro..... 1348  
|||  
11891 ATGATGCGCTTCCGCGTGCAGTGGAGCGAGCTCCAAAGCCCACTCA 11940  
  
1348 ..... 1348  
  
11941 CGCCTGCGCGCGAGCGGCTCTTCTCGGACAGGCGGCGCAGTCT 11990  
  
1349 .....GlyAlaSerProGlyAsnProLeuSer...ProSerLeu 1360  
: : : : :  
11991 CGCGTGCAGCGCGCGCTCGCGGCTAGCGGAGCTCGTGGCTCCGAA 12040  
  
1361 SerAspLysAspArgGlyLeuLys...GlyAlaGlyLysProValGl 1376  
: : : : :  
12041 GCGCCTCGACAGGCGGCTTGGCTCCCGGCTCGTCTGCGCC... 12085  
  
1376 yValGlnGluGlyLeuValAsnValGlyThrGlyLysLeuProThr. 1392  
: : : : :  
12086 ...CCTTCATGATGACCGGAGGAGCTGCTCGGAGGCGCCACGA 12131  
  
1393 .....SerGlyAlaAspProLeuLysArg.....AsnProThrAsn 1404  
: : : : :  
12132 GGCACCGCTCGGCTCGGCTCTTGAAGCTGCTCGCGAGCAAC 12181  
  
1405 ArgSerLeuLysGlyLysLeuMetAsnSerLysLysLeuSerThrAs 1421  
|||  
12182 GCGTGGCTCGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 12231  
  
1421 pCysPheLysThrGlnAlaPheThrSerProGlnAlaLeuGlnProGly 1438  
: : : : :  
12232 CACACGGAAGAGGAGTCAAGGAGCTCG...CTCAGCGCGGCTCTG 12275  
  
1438 lYthAlaLeuAlaProLysLysArgSerArgLysGlyArgAlaGlyAla 1454  
|||  
12276 GGGGCTCGCGCTCGCGCTCGCGCTCGCGCTCGCGCTCGCTCGCT 12325  
  
1455 HisGlyLeuSerLysGlyProLeuGlnLysArgProTyrLeuGlyProAl 1471  
: : : : :  
12326 TCTCGAGATCG.....ACCTACCGAGGCTCCCGAGGCGGCTG 12366  
  
1471 aLeuLeuLeuThrProArgAspArgAlaSerGlyThrGlnGlyAlaSerG 1488  
: : : : :  
12367 CTAGGCGGCTCGACACAGGAGGAGGAGGAGGAGGAGGAGGAGG 12416

1488 luepLysSerGlyGly..... 1493  
: : : : :  
12417 ACCCTCATCCGAGGTGGCAACACCGCTCGACGAGCGCTCATCC 12466  
  
1494 .....GlyGlyLys.....LysProLysMetGluGl 1502  
: : : : :  
12467 CGCGGCAACACCGCTGGCGCTCCATATTCGCAACAAAGGACCTTC 12516  
  
1502 uLeu...GlyProAlaSerGlnProProGlnLysArgProCysGlnProG 1518  
: : : : :  
12517 GACGCGCTGGCTCGCTCGAGCGCGG...AGCCGAGGCGGCT 12560  
  
1518 lnhThrArgAlaGlnProGlyHisThrAsnTyrSerTyrSer 1534  
: : : : :  
12561 CGCACAGCGCAAGTCCGCTCGCTCGAGCGGAGGCTCAACTCC 12610  
  
1535 LysArgLys.....ArgLeuThrArgGlyArgAl 1544  
: : : : :  
12611 GCGATGCTCGACACCGCTTGGCATATTCGCGGCGAGCGCGCGCGC... 12658  
  
1544 aLysAsnThrThrSerSerProCysLysGlyArgAlaLysArgArgG 1561  
: : : : :  
12659 .....TCGGAGCGGAGGCGGCGCATCGTTACTGAAGT 12692  
  
1561 lnhGlnValLeuProLeuAspPro..... 1569  
: : : : :  
12693 CGGTCCAGTGTCTCCCGATACACCGTAGCGGCGGTGATGGGGGTCT 12742  
  
1560 ..AlaGluProGluIleArgLeuLysTyrIleSerSerCysLysArgLe 1585  
: : : : :  
12743 TCGCGCGAGCTTGGTC..... 12760  
  
1585 uArgSerAspSerArgThrProAlaPheSerProPheValArgValGluL 1602  
: : : : :  
12761 ..CCAGGCGATCGCGGAGCGCGCGATGATCT..... 12790  
  
1602 ysArgAspAlaPheThrIleCysThrValValAsnSerProGlyAsp 1618  
: : : : :  
12791 .....GCCCATCCCGCGCGCTGCTCT 12814  
  
1619 AlaProLysProHisArgLysProSer..... 1628  
: : : : :  
12815 TCGCGCAACCGCGAGCGCTCGCTCGCTCGCTCGCTCGCTCGCT 12864  
  
1629 .....SerAlaSerSerSer..... 1633  
: : : : :  
12865 CTGCTGATCTCGGCTCGGCTGAAACCAATCAACGCTGCTCATCCATGC 12914  
  
1634 .....SerSerSerSerSerSerLeuAspAlaAlaGly 1645  
: : : : :  
12915 GCGCGCGCGCGCTCGGAGCGCGCGCTTCTAGCTCGCGCGCGCTCG 12964  
  
1646 AlaSerLeuAlaThrLeuPro.....GlyGlySerIleLeuGlnPro 1659  
: : : : :  
12965 GCGCGAGTCTTTCACCGCGCGCTCGGAGGAGTGGAGCGCTCTCGCG 13014  
  
1659 oArgProSerLeuProLeuSerSerThrMetHisLeuGly...ProVal 1675  
: : : : :  
13015 GCGCTCG.....GCTTCGAGGAGCGCGCGCTCGCGCTCGCTCG 13052  
  
1675 alserLysAla...LeuSerThrSerCys 1683  
: : : : :  
13053 TGACCTGGGCTTCGAGGAGGAGCTTCTCG 13081

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-457-335A-6

seq\_documentation\_block:

; Sequence 6, Application US/08457335A

; Patent No. 5723759

; GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas

; APPLICANT: Ligon, James M.

; APPLICANT: Beck, James Joseph

[illegible]

531 erLeuValLysAspSerSerIysProPheSerLeu..... 643  
:: :: :: :: :: :: :: :: :: :: :: :: ::  
9201 GTCACAGACGCCGATCACACCCCGCAGCTGCAGTGGCTGCC 9250  
544 .....GluAsnHisSerAlaCysLeuAspSerVa 653  
:: :: :: :: :: :: :: :: :: :: :: :: ::  
9251 ATGGCACCAGCACTCTCTGGAGACCCCATCGAGTGC..... 9289  
553 lalalySerAlatrpProArgProGlyLeuProGluLaLeuProAsps 670  
:: :: :: :: :: :: :: :: :: :: :: :: ::  
9290 .....AAGCCCTGGCG...CGCTCTAGCCGACGACAGCCG... 9325  
670 erLeuGlnLeuAspLysGlyGlyAsnAlaLysAspPheSerProGlyLeu 686  
:: :: :: :: :: :: :: :: :: :: :: :: ::  
9326 .....CTGAAGCCCTCTCTCTCGCG 9349  
687 PheGluaspProSerValAlaPheAlaThrProaspProLysLysThrTh 703  
:: :: :: :: :: :: :: :: :: :: :: :: ::  
9350 CGCTCAAGACCAATCGGCCATCTGAGGCCGCTCGGCCCTCGCGGC 9399  
703 rGlyProLeuSerPheGlyThrLysProThrLeuGlyValProAlaProA 720  
:: :: :: :: :: :: :: :: :: :: :: :: ::  
9400 GTCCGCAAGATCG.....TGCCCTCTCCGCCMATGCCCTCGCCCC 9443  
720 spProThrThrAlaAlaPheAspCysPhePro.....AspThr 732  
:: :: :: :: :: :: :: :: :: :: :: :: ::  
9444 CACCCTCCACACGGCCCGCCAAT...CCCTGATTGATTGGGTACA 9489  
733 ThrAlaAlaSerSerAlaAspSerAlaAsnPropheLatrPrGluGl 749  
:: :: :: :: :: :: :: :: :: :: :: :: ::  
9490 CTCGCCATPCAGCTGGTTGATACC.....CGAGGTCTTGGCCCGGCA 9533  
749 uasnLeuGlyAspAlaCysProArgrtpGlyLeuHisProGlyLeuLeu 766  
:: :: :: :: :: :: :: :: :: :: :: :: ::  
9534 CGAA.....GATAGCAGTCCCGC..... 9552  
766 hrLysGlyLeuGlnGlyLysAlaSerAspGlyLysLysGly 782  
:: :: :: :: :: :: :: :: :: :: :: :: ::  
9553 .....CGCGCGAGCTCTCGCCCTCGACTCTCGCGACC 9588  
783 AspThrHis.....GluAlaSerAlaCysLeuGlyPheGlnGl 795  
:: :: :: :: :: :: :: :: :: :: :: :: ::  
9589 AAGCCCAAGCTATCTCGAGAGGCTCCGCCGCCCTG.....TC 9629  
795 uGluaspProGlyGlyLysValAlaSerLeuProGlyAspPheLysG 812  
:: :: :: :: :: :: :: :: :: :: :: :: ::  
9630 GGCGAGCGGCCACCTCACAGAGCGCTCGGACGCTCCCGCGCGGT 9679  
812 lndLuGluValGlyLysGlyGlnAlaGlyLeuLeuGlnCys 828  
:: :: :: :: :: :: :: :: :: :: :: :: ::  
9680 GTGCGGTCTCTGTGCGCCAGGAGCGCGCTCGGCCCGCCAGCG 9729  
829 ProGlu.....ValAlaLysAlaAspargtrPleuGluAs 840  
:: :: :: :: :: :: :: :: :: :: :: :: ::  
9730 AAGCGGCTCGGACACCTCTCGCCACGACGCTCGCCCTTATCGA 9779  
840 pSerArgHisCysSerThr.....AlaaspPhe.GlyAspLeupro 854  
:: :: :: :: :: :: :: :: :: :: :: :: ::  
9780 TGTGCGATTTCGAGGACACACCCCGGCCATCTCGACGACCGCGCG 9829  
985 LeuLeuProThrSerArgLysGluAspLeuGluAlaGluGluTru 871  
:: :: :: :: :: :: :: :: :: :: :: :: ::  
9830 CTCTCTGGCCCGACCGCAGC..... 9853  
871 rSerSerLeuCysGluLeuGlySerProGluInArgrProGlyMetG 888  
:: :: :: :: :: :: :: :: :: :: :: :: ::  
9854 AGCTCTCTCGCGTGCAGCTCGCTCGCCAGGACAAAGCCG..... 9895  
888 lnaSpProLeuSerProLysAlaProteinLecysThrLysGluVal 904  
:: :: :: :: :: :: :: :: :: :: :: :: ::  
9896 .....CCCAGACCGGTCTCGCCGAGCGGCGGCCACG 9931  
905 GluGluValLeuAspSer.....LysAla.....GlyTr 914



10769 TCGCCCTCCGAGAACCTCGAGAGCCCTCACGCGTCATCCGCTGCTGCTC 10818  
1152 hrcInclnLerPhnISerLysArgGlyAlaProSerGlnGlyArgLeu 1168  
10819 GG.....CTCCATTCCAGAGAAAGAAAGCCACTCCG..... 10849  
1189 ProAsnGly.....ArgAl 1173  
10850 CCGCGCTCTCTCTCTCTCTGAGGAGAGCTCTCTACCCGAGGCTCGGCTC 10899  
1173 athrLysLeuLeuAspAsnSerHisLeuProAlaIhrPheLysValS 1190  
10900 GACGTGAGAGAACTTCTTCGAGCGCAACCTCCGAGGAGGTCTGCTCCGCC 10949  
1190 erSerSerProGlnLysGln.....GlyArgValSerGlnArgAla 1203  
10950 CACCTCAACCCCTTCCAGAGAGAGCGGTCTGTGGCTGAGAGCTTCACAGAGC 10999  
1204 ArgVal..... 1205  
11000 AACGCTTCCGAGCTGCTCCGAGGCTCACGAGCTCGGAGGACATCCG 11049  
1206 .....ProLysProGlyAla..... 1210  
11050 CTGTGTGAGGCGCGCGCTGCGCTCTCGCGACAGCGGAGGCTTCTCTTTAC 11099  
1211 .....GlySerLysLeuSerAspArg 1217  
11100 AGGCGGCTCTCTCTCTCGACAGACACCCGTGGCTGAGAGGCATGCGCT 11149  
1218 ProLeuHisAlaLeuLysArgGlySerAlaPhe..... 1228  
11150 TCGGCAACCCCATCTTACCGAGGACCGGAGTTTCTCGAGCTCGCGCTGAC 11199  
1229 .....MetaLarProValProThrLys.....LysArg 1238  
11200 GTCCGCCACCGCGGTGGCGCTCGACACCCGTGCAAGAGCTCAAGCTGAGAGC 11249  
1238 snLeuValLeuArgHisGlySerSerSerSerSerAlaSerAlaMet 1254  
11250 CCGCTTCGCTCTCCCATCGAGAGACGCTCCTCTCCAGAGTCTCCGTCG 11299  
1255 Gly..... 1255  
11300 GGGCCGTGACAGACGACAGAGAGAGGCGCTCTTTCATACGAGCA 11349  
1256 .....GluMetG 1258  
11350 GAGAGCGGCTTCAAGAGAGGCCCTGAGTCCGACGACGAGCGCTCTCT 11399  
1258 LysArgGlyArgGlyLeuArgValProProProSerSerArgGlyCysLeu 1274  
11400 CTCGCGGAGGAGCCCATCCCTCTCCGCCGATCTCCAGAGTGGC..... 11449  
1275 LeuProArg.....LysProSe 1280  
11444 CTCCTCGTAGTGCATCCCGGTGAGACTCGAAGGCTCTACGAGACCTC 11499  
1280 rPro.....ProArgLameAlaSerLeuProG 1290  
11494 GCCAAGCTCGGCTTGCCTTACGCGCCCGAGTTCCAGGGCTTCGCGCTCGT 11549  
1290 InSer.....His 1293  
11544 CTACAAAGCGGAGGAGAGCTCTTTCGCAAGACCAAGCTCCGAGAGCG 11599  
1294 ProArgArg.....ProProPheLeuProGlnAlaIhrGleuSerAlaIhr 1309  
11594 CCGAAAGAGTGGCGCGCGGTTCGCC.....TCCACCTGTGGCGTGCAC 11649  
1309 eGlnGlyAlaMetLysThrLysValLeuProProLysArgGlyArgGly 1326  
11644 AGCGCCCTGTGATGCACAGTGGCTTTGAGAGACGAGAGAGAGAGCGTGGC 11699

1326 eu..... 1326  
11691 TCTGCCCTTCGTGGAGCGAGTCTCGCTCGCTCCGTGGGTGCCACCA 11740  
1327 .....LysLeuGluAlaIleValGlnLysIleThrSerProSer..... 1339  
11741 CTTGGCGGTGGCTTCCACCTGCCAGGAGTGAATCTCGCTTCGATC 11790  
1340 .....LeuLysIlePheA 1344  
11791 GTCTGGCCGAGCGCCGAGGTGACCTCTTGGCTCGGTGCAAGGCTCGC 11840  
1344 lacys..... 1345  
11841 CATTGGGAGACGTCCGCCGCCAGCTCCGACCCCGGAGCTTCCACCC 11890  
1346 .....LysAlaPro..... 1348  
11891 ATGATGCGCTCTTCGGGTGACTGGAGCGAGCTCCAAAGCCCACTTCA 11940  
1348 ..... 1348  
11941 CGCCTGCCGCCGCGAGCGGCTCTTCTCGGCACAGCGGCCACGATCT 11990  
1349 .....GlyAlaSerProGlyAsnProLeuSer...ProSerLeu 1360  
11991 CGCGCTCACGCCCGCTCGCCGCTAGCGGACCTCGCTGCCCTCCGAA 12040  
1361 SerAspLysAspArgLysLeuLys...GlyAlaGlySerProValG 1376  
12041 GCGCCTCGACAGGCGGTGCTGCTCCGCGCTCGCTCGCTCGCC.... 12085  
1376 yValGluGluLeuValAsnValGlyThrGlyGlnLysLeuProThr. 1392  
12086 ...CCTTCATCATGACCGAGCGAGCTCGTCCGAGCGCCACGA 12131  
1393 .....SerGlyAlaAspProLeuLysArg.....AsnProThrAsn 1404  
12132 GGCACCGCGCTCGACTCGCTCTTCGACGCTCGCTCGCGGACGAC 12181  
1405 ArgSerLeuLysGlyLysLeuMetAsnSerLysLysLeuSerThrAs 1421  
12182 GCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 12231  
1421 pCysPheLysThrCluAlaPheThrSerProGluAlaLeuGlnProGly 1438  
12232 CACACGAGAGCAGCTGACGAGGACCTG.....CTCAGCGCGCTCTG 12275  
1438 lyThrAlaLeuAlaProLysLysArgSerArgLysGlyArgAlaGlyAla 1454  
12276 GGGCTCGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 12325  
1455 HisGlyLeuSerLysGlyProLeuGluLysArgProTyLeuGlyProAl 1471  
12326 TGTCTGACATCG.....ACCTCAGAGGCGCTCCGACGAGCGCTG 12366  
1471 alaLeuLeuThrProArgAspArgAlaSerGlyThrGlnGlyAlaSerG 1488  
12367 CTAGCGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 12416  
1488 luAspAsnSerGlyGly..... 1493  
12417 ACCCTCATCCGAGGTGGCGGACACGCTGAGCGAGCGCTCATCC 12466  
1494 .....GlyGlyLys.....LysProLysMetGluG 1502  
12467 CGCGGACGACCCACCTGCGCTCCATATTCGACGAGGACCTTC 12516  
1502 uLeu...GlyProLaserGlnProProGluGlyArgProCysGlnProG 1518  
12517 GAGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 12560

1538 lnThrArgAlaGlnLysGlnProGlyHisThrAsnTySerSerTySer 1534  
12561 CGCACGCGCCAAAGTCCGCGATCCCGCTCGACGCGGAGGCTCAACTCC 12610  
1585 LysArgLys.....ArgLeuThrArgGlyArgAl 1544  
12611 CGCATCTCGCTCGACACCTTGGCATGTATTCGGGCGAGCGCGCGCGC... 12658  
1544 aLysAsnThrThrSerSerProCysLysGlyArgAlaLysArgArgG 1561  
12659 .....TCGGAGCGGAAGCGCGGCGCATCGTTACTGAAGT 12692  
1561 lngLingLynValLeuProLeuAspPro..... 1569  
12693 CGTCCAGGTGTCTCCGATACACCTAGGCGACCGGGGTGATGGGGGTCT 12742  
1570 .....AlaGluProGluIleArgLeuLysTyThrIleSerSerCysLysArgLe 1585  
12743 TCGCGCAGCCTTTGGTC..... 12760  
1585 uArgSerAspSerArgThrProAlaPheSerPropheValArgValGlu 1602  
12761 CCACGCGCATCGCGCGACGCGCGCATGATCT..... 12790  
1602 ysArgAspAlaPheThrIleCysThrValValAsnSerProGlyAsp 1618  
12791 .....GCCCCATCCCCCGCGCTGTCT 12814  
1619 AlaProLysProHisArgLysProSerSer..... 1628  
12815 TCGCCACGCGCCAGCGCTCCCATCATCTATCTACCGCTACTATGGA 12864  
1629 .....SerAlaSerSerSer..... 1633  
12865 CTGCTCGATCTCGGCGATCTGAACCCATCAACGTCTCTCATCCATGC 12914  
1634 .....SerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 1645  
12915 GCGCGCGCGCGCTCGGCGCGCGCGCTTCTAGCTCGACGCGACCTCG 12964  
1646 AlaSerLeuAlaThrLeuPro.....GlyGlySerIleLeuGlnPr 1659  
12965 GCGCGAGGTCTTTCGCCAGCGAGTCCAGGGAAGTGGAGCGCTCTCGCG 13014  
1659 oArgProSerLeuProLeuSerSerThrMetHisLeuGly...ProVal 1675  
13015 GCGCTCG.....GTTTCGACGATCGGACCTCGCTCGCTCGCTCG 13052  
1675 alSerLysAla...LeuSerThrSerCys 1683  
13053 TGACCTGGCTTCGAGCAGCAGCTTCTGCG 13081

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-729-214-6

seq\_documentation\_block:

; Sequence 6, Application US/08729214

; Patent No. 5817502

; GENERAL INFORMATION:

; APPLICANT: Ligon, James M.

; APPLICANT: Hill, Dwight Steven

; APPLICANT: Ryals, John Andrew

; APPLICANT: Hammer, Phillip E.

; APPLICANT: van Pee, Karl-Heinz

; APPLICANT: Kirner, Sabine

; TITLE OF INVENTION: Genes for the synthesis of

; antipathogenic substances

; NUMBER OF SEQUENCES: 27

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Ciba-Geigy Corporation

; STREET: 520 White Plains Road

; CITY: Tarrytown

; STATE: NY

; COUNTRY: USA



470 rArgThrProGluGlnHnHisLysSerGlnHnHisCysSerProGluGlySerg 487  
: : ||||| : || : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :  
8590 CTCAGCCCC.....ACGACCCTCGTCC.....TCGA 8618

487 lYtyr-SerAlaGluProAlaGlyThrProLeuSerGluProPro..... 501  
: : :::::||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| :  
8619 ATCTGCCTGGCAGGCCCTCGAAGACGGCGGCATGTCCCCCCCACCTCA 8668

502 .....SerSerThrProGlnSerThrHI 509  
: : :::::||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| :  
8669 AGGATTCOCCACCGCGCTCTTCGTGGCATCGGCGCACGAATACGCA 8718

509 sAlaGlu.....ProGlnGlnAla 516  
: : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :  
8719 TTGCGAGGAGCGACACGAGAATCGAGCTTAGCCTTAACCTTCACAGGCA. 8767

516 spTyrlleuSerGlySerGluAspProLeuGluArgSerPheIeuTyrcys 532  
: : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :  
8768 .....CGCGGGGCTCTTTGCCGGGGCGCTTGCCCTACA 8803

533 AsnGlnAlaArgGlySerProAlaArgValAsnSerIasnSerIyAlaLy 549  
: : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :  
8804 CGTCTGGCTCGAAGGGCCG...CGCTCTCGGTGCACCGCTCGCTCC 8850

549 sProGluSerValSerThrCysSerValThrSerProAspAspMetSart 566  
: : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :  
8851 TCCTGCCTGTGCGCTCTCACTCGCTGCCAGGCCCTCGACACGGGCGA 8900

566 hrlYserAspSerPheGlnSerLeuHisGlySerLeuProLeuAsp 582  
: : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :  
8901 GTGCAACCTCGCCCTCGCGGGCGTCTCGTCTATGGCTCCCGCGAGG 8950

583 SerPheSerLysPheValala..... 589  
: : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :  
8951 GCTGTGCTCTCTTCTCCGCTCGCGCTGTGGCGCCGACGGCGCTCC 9000

590 .....GlyGluArgAspCyp 595  
: : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :  
9001 ANGACTTCTCGGCCAACCGCAGCGCTACGGACCGGAGAGGGCTCAT 9050

595 roArgLeuLeuSerAlaLeu.....AlaGlnGluAspLeuAla 608  
: : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :  
9051 CGTCTTCGCTCGAGCGCTCGGTGAGCGCCCTGTGGCGCCGACGGCGCTCC 9100

609 serGluIleLeuGlnGluAlaIleGlyGluIylsAlaAspLysAl 625  
: : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :  
9101 TCTTCGCTCGTCCGGCACCAGCATCAACACGACGGCGGTGAGC 9150

625 a.....TrpAlaGluAlaProS 631  
: : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :  
9151 GGATATCACGCCGCCAACCGCACCTCCACAGAAAGGTCTCGCGCGCG 9200

631 erLeuValIylsAspSerSerLysProPhoPheSerLeu..... 643  
: : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :  
9201 GCTCCACGACGCGCGATCACCCCGCGGACTGTGAGCTCGTGTGAGTGC 9250

644 .....GluAsnHisSerAlaCysLeuAspSerVa 653  
: : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :  
9251 ATGGCACC GCCACTCTCTTGGAGACCCCATCGAGTGC..... 9289

653 lAlaLeuSerAlaIrrProArgProGlyGluProGluAlaLeuProAsps 670  
: : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :  
9290 .....AAGCCCTGGCG...CGCTCTACGGCAGCAGCACCCG.... 9325

670 erLeuGlnLeuAspLysGlyIylasnaLaIylsAspPheSerPheLeu 686  
: : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :  
9326 .....CTGAAGAGCTCTCTCTCTCGCG 9349

687 PheGluAspProSerValAlaPheAlaThrProAppProIylsTh 703  
: : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :  
9350 CGGTCAAGACCAATCGGCCCATCTCGAGCGCGCTCGCGCTCGCGGCG 9399

703 rglyProIeuSerPheGlyThrIysProThrLeuGlyValProAlaProA 720  
9490 GTCCGAAGATCG.....TCGCCTCCTCCGCCATGACGGCCTCCGCC 9443  
720 spProThrThrAlaIlePheAspCysPhePro.....AspThr 732  
9444 CACCCTTCCACACGGCGCCGGCGCAT.....CCTTGATGTATTGGATACA 9489  
733-ThrAlaIleSerAlaAspSerAlaAsnProPheAlaTrpProGluGl 749  
9490 CTCGCCATCGACGCTGTTGATACC.....CGAGGTCCTGGGCCCGCCA 9533  
749 uasLeuGlyAspAlaCysProArgTrpGlyLeuHisProGlyGluLeu 766  
9534 CGAA.....GATACGACTCCCCG..... 9552  
766 hrLysGlyLeuGluGlnGlyGlyLysAlaSeraspGlyIleSerLysGly 782  
9553 .....CGCGCGCGGCTTCGGCTTCGGACTCTCCGGCAC 9588  
783 AspThrHis.....GluAlaSerAlaCysLeuGlyPheGlnGl 795  
9589 AAGCCCAAGTCATCTCGAGGAGGCTCCCGCGCGCTG.....TC 9629  
795 uGluAspProGlyGluLysValaSerLeuProGlyAspPheLysG 812  
9630 GGGCGAGCGCGCCACTCACAGAGCGCTCGCGACCGCTCCCGCGCGGT 9679  
812 lnglucLuValglGlyValLysGluGlnAlaGlyGlyLeuLeuGlnCys 828  
9680 GTCCGTGCTCTGTCGCCAGAGGAGCGCGCGCTCGCGCCACGCG 9729  
829 ProGlu.....ValAlaLysAlaArgTrpLeuGluAs 840  
9730 AACGGCTCCGCGACCACTCTCGCCACGACGACTCGCCCTATCGA 9779  
840 pSerArgHisCysCysSerThr.....AlaAspPhe.GlyAspLeuPro 854  
9780 TGTGGCTATTGCGAGGCCACCAACCGCGCCACTTCGAGCACCGCGCG 9829  
855 LeuLeuProThrSerArgLysGluAspleuGluAlaGluGluLutry 871  
9830 CTCTCTCGCGCGAGCGCGAGCG..... 9853  
871 rSerSerLeuCysGluLeuLeuGlySerProGluGlnArgProGlyMetG 888  
9854 AGCTCTCTCGCGCTCGACTCGCTCGCCACGACGACGCG..... 9895  
888 lnAspProLeuSerProIysAlaProLeuilleCysThrLysGluGluVal 904  
9896 .....CCCCGAGCACCGTTCTCGCGCGGAGCGGAGCCAGC 9931  
905 GluGluValLeuAspSer.....Lysala.....GlyTr 914  
9932 GCAAGCTCTCTCTGCTCTTCTCGCGCACGAGCTCGCAGTGGGAGGATG 9981  
914 pGlySerProCysHisLeuSerGlyGluSerValIleLeuLeuGlyProt 931  
9982 CCCTCTCCCTGCTGACTCTCGCGGCTTCCTCGCGCTCAGCTCGAGC 10031  
9931 hrValGlyThrGluSerLysValInSerTrpPheGluSerSerLeuSer 947  
10032 ATCGAGCGCGGCTCGCTCTCAGCTG..... 10060  
948 HisMetLysProGlyGluGluGlyProAsp.....GlyGluArgAlaPr 962  
10061 .....AGTGGAGCTGCTCGCGCTCTCGCGCGGAGGCGCC 10101  
962 oGlyAspSerThrThrSerIspla..... 970  
10102 CCTCTCTCGACCGCGTCTGAGCTGCTGACACCGCGCTCTTTCGGCTAT 10151  
971 .....SerLeuAlaGlnLysProASNLSyProAlaVal 981

```
10152 GGTCTCCCTGCGCCGCTCTGCGCTGCGCTGAGCGGCGCGCG 10201
982 ProGluAlaProIleAlaLysGluProValProArgLysSerle 998
10202 TCGTCGCGCAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10251
998 uArgSerArgValHisArgLysGluProGluAlaGluAspSerPro 1015
10252 CTCTCCCTGAGGAGCGGCGCGCGCGCGCGCGCGCGCGCGCG 10286
1015 ys.....ArgAlaProValLeuProLysAspLeuLeuProGlu 1028
10287 GCGCAGGAGAGGCGCTCACACCGTCGCGCGCACGCGCGCGCGCG 10336
1029 SerCysThrGlyProProGluGlyGln.....MetGluVala.. 1041
10337 TCGAGCTCGGCGGCTCGGAGCTCCAGACCTACCTCGCTCGGCGC 10386
1042 .....GlyAlaProGlyArgGlyAlaSerGluG 1051
10387 AGGCTCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10436
1051 LysLeuProArgMetCysThrArgSerLeuThrAlaLeuSerGluProArg 1067
10437 CGAGCCCGCGCGCGCTCGAGCGCGCTGCTCAGCGCGCGCGCG 10486
1068 ...ThrProGlyProProGlyLeuThrThrProAlaPro..... 1080
10487 TGTTCGCGCGCAGATCCGCGTGAATACGCTCCAGCTCCGCGC 10536
1081 .....ProAspLysLeuGlyG 1086
10537 GAGCGCTCCAGAGACGAGCTCGCGAGGCTCAGCAACATCGCT 10586
1086 LysGluAlaArgAlaAlaPheLysSerGlyLysArgValGlyLysProSer 1102
10587 GAGCTGCGAGCTCCCTCTTATT.....CGACGCTCA 10618
1103 ProLysAlaAlaSer...SerProSerAspProAla..... 1113
10619 CCGGACACAGGCTCGAGCGCTCGAGCTCGAGCGCGGCTACTG 10668
1114 .....AlaLeuProValAlaSerAspSerProMetGlySerLysT 1128
10669 AACCTCCGCGCAACCGCTCTGCTTCGAGCGGACGAGCGGCT 10718
1128 hrLysGluThrAspSerProSer..... 1135
10719 CGATGGGCACTGCTTCTCGTCAAGGTCAGCCCGCCATCCGCT 10768
1136 ThrProGlyLysAspGlnArgSerMetIleLeuArgSerArgThrLysT 1152
10769 TCGGCTCGCGGAGACGCTCGAGCGCTCAACGCTCGATCCCG 10818
1132 hrGlnGluIlePheHisSerLysArgArgArgProSerGluGlyArgLeu 1168
10819 GG.....CTCCATTCGAGCAGAGAGAGGCGCACCTCG..... 10849
1169 ProAsnCys.....ArgAl 1173
10850 CCGGCTGCTCTCTCTCTGCGGCGAGCTCTACCCGAGGCTCGG 10899
1173 arThrLysLysLeuLeuAspAsnSerHisLeuProAlaThrPheLysValS 1190
10900 GACTGAGAGAGACTTCTTCGCGCCCTACGCTCCCGCAAGTCTCC 10949
1130 erSerSerProGluLysGlu.....GlyArgValSerGlnArgAla 1203
10950 CACCTACCGCTTCAGCAGAGAGGCTTCGCTGAGCTCCACAGC 10999
1204 ArgVal..... 1205

11000 AACGCTTCGAGAGCTGCTCTCCGCAAGCTTACCTGGCGCGCAATCCG 11049
1206 .....ProLysProGlyAla..... 1210
11050 CAGCTCGGCGCGCGCGCTGCGCTTCGCGAGCGGCTTCCTTC 11099
1211 .....GlySerLysLeuSerArg 1217
11100 AAGGCGGCTCTCCCTCGCAGAGCACCCGCTCGAAGGCGGAGCG 11149
1218 ProLeuHisAlaLeuLysArgLysSerAlaPhe..... 1228
11150 TCGGCACACCACTTACCGGCGACGCGCTTCTGAGCTCGCCCTG 11199
1229 .....MetAlaProValProThrLys.....LysArg 1238
11200 GTCGCGCACCGCGCTGCGCTGAGCAGCGTGAAGAGCTCAGCG 11249
1238 snLeuValLeuArgHisGlySerSerSerSerSerAlaSerAlaMet 1254
11250 CCTCTGCTCTCTCCATTCGACAGGACCGCTCTCTCCATCTCG 11299
1255 Gly..... 1255
11300 GGCCTGTGACGACGAGAGAGAGAGGCGCTCTTCTCCATAGCC 11349
1256 .....GluMet 1258
11350 GAGGAGCGGCTGAGAGTGGCCCGGAGACTCGCAGCGAGCGGCT 11399
1258 LysArgArgArgGlyLeuArgValProProProSerSerArgLysLeu 1274
11400 CTCGCGCGCGAGCCCATCCCTCTCGCGGATCTCCAGATCG 11443
1275 LeuProArg.....LysProse 1280
11444 CTCCCTCGAGTGCATCCCGGTGAGCTCGAAGGCTCTACGACCT 11493
1280 rPro.....ProArgAlaMetAlaSerLeuProG 1290
11494 GCCAACCTCGGCGCTGCTACGCGCCGAGATTCAGAGGCTCCG 11543
1290 InSerSer.....His 1293
11544 CTACAAAGCGCGGAGAGCTTGTGCCGAAGCAAGCTCCCGAG 11593
1294 ProArgArg...ProProPheLeuProGluAlaArgLeuSerAlaAlaP 1309
11594 CCGAAGAGATGCCGCGCGGTTGCC...TCCACCTTCGCTCTGAC 11640
1309 eGlnGluValaMetLysThrLysValLeuProProArgLysGlyArgLys 1326
11641 AAGGCGCTTCGAGCTGCTTGAAGAGAGCAGACAGAGAGGAG 11690
1326 eu..... 1326
11691 TCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 11740
1327 .....LysLeuGluAlaIleValAlaLysLysSerProSer..... 1339
11741 CCTGCGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 11790
1340 .....LeuLysLysPhe 1344
11791 GTCGCGGAGCGCGCGAGTACAGCTTGTCTCTCTCTCTCTCT 11840
1344 LacyS..... 1345
11841 CATGGAGGAGAGCTCGCGCGCGAGCTCGGAGCCCGAGCTCC 11890
1346 .....LysAlaPro..... 1348
11891 ATGATGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 11940
```

1348 ..... 1348  
11941 CGCGCTGCGCGCGCGCGCGCTCTTCGCGACAGCGCGCCACGATCT 11990  
1349 .....GlyAlaSerProGlyAsnProLeuSer...ProSerLeu 1360  
11991 CGCGCTGCGCGCGCGCGCTGCGCGTACGCGCGCTGCTGCTCGGAA 12040  
1361 SerAspHisArgGlyLeuLys...GlyAlaGlySerProValGI 1376  
12041 GCGCGCTGCGCGCGCGCTGCGCGTACGCGCGCTGCTGCTGCGCG 12085  
1376 yValGluGluLeuValAsnValGlyThrGlyGlnLysLeuProThr. 1392  
12086 ....CCTTCATCATGCGCGCGCGCGCTGCTGCGCGCGCGCGCG 12131  
1393 .....SerGlyAlaAspProLeuCysArg.....AsnProThrAsn 1404  
12132 GCGCGCGCGCTGCGCTGCGCGCTGCTGCGCGCTGCGCGCGCGCG 12181  
1405 ArgSerLeuLysGlyLysLeuMetAsnSerLysLysLeuSerThrAs 1421  
12182 GCGCGCTGCGCGCGCTGCGCGTCCGCGTACCGCGCGCGCGCGCG 12231  
1421 pCysPheLysThrGluAlaPheThrSerProGluAlaLeuGlnProGly 1438  
12232 CACCGCGCGCGCTGCGCGTCCGCGTACCGCGCGCGCGCGCGCG 12275  
1438 lYThrAlaLeuAlaProLysLysArgSerArgLysGlyAlaGlyAla 1454  
12276 GGGGCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12325  
1455 HisGlyLeuSerLysGlyProLysGlyArgProTyLeuGlyProAl 1471  
12326 TGCTGCGATCG.....ACCTCAGCGCGCGCGCGCGCGCGCG 12366  
1471 aLeuLeuLeuThrProArgAspArgAlaSerGlyThrGlnGlyAlaSer 1488  
12367 CTAGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12416  
1488 luAspAsnSerGlyGly..... 1493  
12417 ACCCTCATCCGAGGTGGCGACCGCGCGCGCGCGCGCGCGCG 12466  
1494 .....GlyGlyLys.....LysProLysMetGluGI 1502  
12467 CGCGCGACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12516  
1502 uLeu...GlyProAlaSerGlnProGluGlyArgProCysGlnProG 1518  
12517 GAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12560  
1518 lNThrArgAlaGlnLysGlnProGlyHisThrAsnTySerSerTySer 1534  
12561 CGCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12610  
1535 LysArgLys.....ArgLeuThrArgGlyArgAl 1544  
12611 GCGATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12658  
1544 aLysAsnThrThrSerProCysLysGlyAlaLysArgArgArg 1561  
12659 .....TCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12692  
1561 lngLngLngValLeuProLeuAspPro..... 1569  
12693 CGGTCCAGGTGCTCCCGATACACCGTAGGCGCGCGCGCGCG 12742  
1570 ...AlaGluProGluLeuArgLeuLysTyrlleserSerCysLysArgLe 1585  
12743 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12760

1585 uArgSerAspSerArgThrProAlaPheSerProPheValArgValGluL 1602  
12761 ..CGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12790  
1602 ysArgAspAlaPheThrThrIleCysThrValValAsnSerProGlyAsp 1618  
12791 .....GCCCATCCCGCGCGCGCGCGCGCGCGCGCGCG 12814  
1619 AlaProLysProHisArgLysProSer..... 1628  
12815 TCGCCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12864  
1629 .....SerAlaSerSerSer..... 1633  
12855 CTGCTCGATCTCGGCGCATCTGAACCCATCATCGTGTCTCATCCATCG 12914  
1634 .....SerSerSerSerSerSerSerSerSerSerSerSer 1645  
12915 GCG 12964  
1646 AlaSerLeuAlaThrLeuPro.....GlyGlySerIleLeuGlnPr 1659  
12965 GCGCGAGGTCTTTGCCACCGCGCGCGCGCGCGCGCGCGCG 13014  
1659 oArgProSerLeuProLeuSerSerThrMetHisLeuGly...ProVal 1675  
13015 CGCGCTG.....GCTTCGACGATCGCGCGCGCGCGCGCG 13052  
1675 alserLysAla...LeuSerThrSerCys 1683  
13053 TGACCTGGGCTTCGAGCAGCGCGCGCGCGCGCGCGCGCG 13081

seq\_name: /cgn2\_6/ptodata/2/lna/6A\_COMB.seq:US-09-028-934-6  
seq\_documentation\_block:  
; Sequence 6, Application US/09028934  
; Patent No. 6117670  
; GENERAL INFORMATION:  
; APPLICANT: Ligon, James M.  
; APPLICANT: Hill, Dwight S.  
; APPLICANT: Lam, Steven T.  
; APPLICANT: Hammer, Phillip E.  
; APPLICANT: van Pee, Karl-Heinz  
; APPLICANT: Kirner, Sabine  
; APPLICANT: Young, Thomas R.  
; TITLE OF INVENTION: Pyrrolinrin Biosynthesis Genes and Uses  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6117670artis Corporation  
; STREET: 3054 Cornwallis Road  
; CITY: Research Triangle Park  
; STATE: NC  
; COUNTRY: USA  
; ZIP: 27709  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/028,934  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/729,214  
; FILING DATE: 09-OCT-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/258,261  
; FILING DATE: 08-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meigs, J. Timothy  
; REGISTRATION NUMBER: 38,241



509 sAlaGlu.....ProGlnGluAla 516  
|||||  
8719 TTGGAGAGGAGGAGCAGCAGAGATTCGACCTTATGCTCCATCAAGCA. 8767  
516 spYrLeuSerGlySerGluAspProLeuGluArgSerPheLeuTyrCys 532  
|||||  
8768 .....CCGCGGGTCTTTCCGCGGGGCGCTGGCGCTACA 8803  
533 AsnGlnAlaArgGlySerProAlaArgValAsnSerAsnSerLysAlaLys 549  
:|||||  
8804 CGTCTGGCTGCAAGGCGCG...CGCTCTGGTGACACCGCTGTGCTC 8850  
549 sProGlnSerValSerThrCysSerValThrSerProAspAspMetSerT 566  
|||||  
8851 TCTCTCTCTGCGCTCCACCTTCGCTCGCTCCGACGCTCCGACAGGCGCA 8900  
566 hrLysSerAspSerPheGlnSerLeuHisGlySerLeuProLeuAsp 582  
:|||||  
8901 GTGCAACCTCGCTCGCGGGGCTCTCGCTCATGGCTCCCGCGAGG 8950  
583 SerPheSerLysPheValAla..... 589  
8951 GCTTCGCTCTCTTCCTTCGCGCTTCGCGGCTTGGCGCGCAGCGCGCTCC 9000  
590 .....GlyGluArgAspCysP 595  
9001 AAGACCTTCGCGCAACGCGGCTACGACGCGGAGAGGCGCTCAT 9050  
595 roArgLeuLeuSerAlaLeu.....AlaGlnGluAspLeuAla 608  
|||  
9051 CGTCTCTGCTTCGAGGCGCTGGTACGCGCTCGCGCGGAGCAGCAGCG 9100  
609 SerGluLeuLeuGlnGlnAlaIleGlyLeuLysAlaAspLysAl 625  
:|||||  
9101 TCTTCGCTCTCGCGGCGCAGCGCTCAACAGCAGCGCGCTCGAGC 9150  
625 a.....TrpAlaGluAlaPro 631  
9151 GGTATCAGCGCGCGCAACGCGCCTCCAGCAGAGGTCTCTCGCGCGC 9200  
631 erLeuValLysAspSerSerLysProProPheSerLeu..... 643  
9201 GCTCCAGCAGCGCGCGATCACCGCGCGCGAGCTCGAGTCGTCGAGTGC 9250  
644 .....GluAsnHisSerAlaCysLeuAspSerVa 653  
9251 ATGGCAGCGCGCCTCTCTGGGAGACCCATCGAGTGC..... 9289  
653 lAlaLysSerAlaTrpProArgProGlyGluProGluAlaLeuProAspS 670  
:|||||  
9290 .....AAGCCTCGCGG...CGCTCTACGCGCAGCGCAGCAGCG... 9325  
670 erLeuGlnLeuAspLysGlyGlyAsnAlaLysAspPheSerProGlyLeu 686  
9326 .....CTGAAAGCCTCTCTCTCTCGCGC 9349  
687 PheGluAspProSerValAlaPheAlaThrProAspProGlyLysThrTh 703  
:|||||  
9350 CGCTCAAGACCAACATCGCGCATCTCGAGCGCGCTCTCGCGCTCGCGGC 9399  
703 xGlyProLeuSerPheGlyThrLysProThrLeuGlyValProAlaProA 720  
:|||||  
9400 GTGCCCAAGATCG.....TCGCGCTCTCTCGCGCATGACGCGCTCGCGC 9443  
720 spProThrThrAlaAlaPheAspCysPhePro.....AspThr 732  
:|||||  
9444 CACCTTCACAGCGCGCGCGCAT...CCCTTGATTGATTGGATACA 9489  
733 ThrAlaAlaSerAlaAspSerAlaAsnProPheAlaTrpProGlnG 749  
:|||||  
9490 CTCGCCATCGAGCTGTTGATACC.....CCGAGGTCTTGGCGCGCGCA 9533  
749 uAsnLeuGlyAspAlaCysProArgTrpGlyLeuHisProGlyGluLeuT 766

9534 CGAA.....GATAGCAGTCCCGC..... 9552  
766 hrLysGlyLeuGlnGlnGlyLysAlaSerAspGlyLysSerLysGly 782  
:|||||  
9553 .....CGCGCGGGCTCTCGCTTCGGACTCTCCGGCACC 9588  
788 AspThrHis.....GluAlaSerAlaCysLeuGlyPheGlnG 795  
:|||||  
9589 AAGCCACAGTCTCTCTCGAGGAGCTCCGCGCGCTG.....TC 9629  
795 uGluAspProProGlyGlnLysValAlaSerLeuProGlyAspPheLysG 812  
:|||||  
9630 GGGCAGCGCGCGCCTCACAGAGCGCTCGCAGCCTCCCGCGCGCT 9679  
812 lGlnGluValGlyValLysGlnGlnAlaGlyGlyLeuLysGlnCys 828  
:|||||  
9680 GTGCGCTCTCTCTCGCGCAGGAGCGCGCTCCGCGCGCGCGCG 9729  
828 ProGlu.....ValAlaLysAlaAspArgTrpLeuGluAs 840  
:|||||  
9730 AAGCGCTCCGCGACCTCTCTCGCGCAGCAGCCTCGCGCTTATCGA 9779  
840 pSerArgHisCysCysSerThr.....AlaAspPhe.GlyAspLeuPro 854  
:|||||  
9780 TGTGGCTTATTCGAGCGCACCACCGCGCGCTTCGAGCAGCGCGCG 9829  
855 LeuLeuProProThrSerArgLysGluAspLeuAlaGlnGluLys 871  
:|||||  
9830 CTCTCTCGCGCGCGCGCGAG..... 9853  
871 rSerSerLeuCysGlnLeuLeuLysSerProGlnGlnArgProGlyMetG 888  
:|||||  
9854 AGCTCTCTCGCGCTCGACTCGCTCGCGCGCGCGCAGCAGCGCG..... 9895  
888 lAspProLeuSerProLysAlaProLeuLysCysThrLysGlnGluVal 904  
:|||||  
9896 .....CCGCGCAGCAGCTCTCTCGCGCGCGCGAGCAGCGCAG 9931  
905 GluGluValLeuAspSer.....LysAla.....GlyTr 914  
:|||||  
9932 GCAGGTCTCTCTCTCTCTCGGCAAGCTCGCAGTGGGAGGAGGATG 9981  
914 pGlySerProCysHisLeuSerGlyGluSerValLeuLeuLysGlyProT 931  
:|||||  
9982 GCCT 10031  
931 hrValGlyThrGlnSerLysValGlnSerTrpPheGlnSerSerLeuSer 947  
:|||||  
10032 ATGCGAGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10060  
948 HisMetLysProGlyGlnGlnGlyProAsp.....GlyGluArgAlaPr 962  
:|||||  
10061 .....AGTGGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10101  
962 oGlyAspSerThrThrSerAspAla..... 970  
:|||||  
10102 CCT 10151  
971 .....SerLeuAlaGlnLysProAsnLysProAlaVal 981  
10152 GGTCT 10201  
982 ProGluAlaProLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 998  
:|||||  
10202 TCGTCT 10251  
998 uArgSerArgValHisArgGlyLeuProGluAlaGluAspSerProC 1015  
:|||||  
10252 CT 10286  
1015 ys.....ArgAlaProValLeuProLysAspLeuLeuProGlu 1028  
:|||||



10287 GCGAGGAAAGCGCTCACACGCTGCGGCGCAAGCGGCGATGCGCGG 10336  
1029 SerCysThrGlyProProGlnGlyLeu.....MetGlnGlyAla.. 1041  
10337 TCGAGCTCGGCGCTCGACCTCGACCTACCTCGCTCGCGGGGCGAC 10386  
1042 .....GlyAlaProGlyArgGlyAlaSerGln 1051  
10387 AGGCTCTCACGCGCGCGCTCACAGGCCCGAGGCTACCTCGCTACCGG 10436  
1051 LysLeuProArgMetCysThrArgSerLeuThrAlaLeuSerGlnProArg 1067  
10437 CGAGCCCGCGCGCGCTCGACGCTCGACGCTCGACGCGACCAAG 10486  
1068 .....ThrProGlyProProGlyLeuThrThrProAlaPro..... 1080  
10487 TGTTCGCGCGCAAGTCCGCTGAGTACGCTCGACCTCGACCTCGACATG 10536  
1081 .....ProAspLysLeuGly 1086  
10537 GAGCGCTCGCAAGAGAGCTCGCGGAGGCTACGCAACATGCTCTCG 10586  
1086 LysGlnArgAlaAlaPheLysSerGlyLysArgValGlyLysProSer 1102  
10587 GAGCTCGAGCTCGCTCTTAT.....CGACGCTCA 10618  
1103 ProLysAlaAlaSer.....SerProSerAsnProAla..... 1113  
10619 CGGCGACAGAGCTCGAGGCTCGAGCTCGAGCGGAGGATCGATCGA 10668  
1114 .....AlaLeuProValAlaSerAspSerProMetGlySerLys 1128  
10669 AACCTCGGCAACCGCTCTGCTCGAGCGAGCGAGGCTCTCTCGA 10718  
1128 hTrpGlnThrAspSerProSer..... 1135  
10719 CGATGGCACTCGCTCTCGGCGAGGTCAGCCCGATCCGCTCGACGC 10768  
1136 ThrProGlyLysAspGlnArgSerMetLysLeuArgSerArgThrLys 1152  
10769 TCGGCTCGCGAGAGCTCGGAGCGCTCGACGCTCGATCGCTCGCTCG 10818  
1152 hTrpGlnThrPheHisSerLysArgArgArgProSerGlnGlyArgLeu 1168  
10819 GG.....CTCATTCGACGAGAGAGGCGACCTCG..... 10849  
1169 ProAsnCys.....ArgAla 1173  
10850 CCGGCTCGCTCTCTCGGCGAGGCTCGACCTCGAGGCTCGGCTCG 10899  
1173 aThrLysLysLeuAspAsnSerHisLeuProAlaThrPheLysVal 1190  
10900 GACTGAGAGAGCTCTCGGCGCTCGAGCTCGCGAGGCTCGCTCG 10949  
1190 eTrpSerProGlnLysGlu.....GlyArgValSerGlnArgAla 1203  
10950 CACCTACCCCTTCAGCGAGAGGCTGCTCGAGCTCGACGAGGAGC 10999  
1204 ArgVal..... 1205  
11000 AAGGCTTCGAGAGCTCGCTCGGAGGCTCGACCTCGGCGGACCAATCCG 11049  
1206 .....ProLysProGlyAla..... 1210  
11050 CTGCTCGGCGCGCGCTCGCTCGCGGAGCGGCTCGCTCTCTTAC 11099  
1211 .....GlySerLysLeuSerAspArg 1217  
11100 AGGCGGCTCTCGCTCGAGAGAGCCCGGCTCGAGAGGCGCATGCCGCT 11149  
1218 ProLeuHisAlaLeuLysArgLysSerAlaPhe..... 1228  
11150 TCGGCAACCCATCTCTACCGGCGACCGGCTTCGAGCTCGCTCGAC 11199  
1229 .....MetAlaProValProThrLys.....LysArg 1238  
11200 GTGCGCCACCGGCTCGGCTCGACACCGCGCAAGAGCTCGAGGCG 11249  
1238 snLeuValLeuArgHisGlySerSerSerSerSerAlaMet 1254  
11250 CCGCTCGGCTCTCGCATTCGAGAGAGAGCGCTCTCTCTCGAGCTCGCTCG 11299  
1255 Gly..... 1255  
11300 GGCGCGGAGCAGCAGAGAGAGGCGCTCTCTTCATACCGCAAA 11349  
1256 .....GluMet 1258  
11350 GAGAGCGCTTCAGATGCGCCCTGAGCTGCGACCGACGCGCTCT 11399  
1258 LysArgArgArgLysLeuArgValProProProSerSerArgLysLeu 1274  
11400 CTGCGCGGAGACCCATCCCTCTCGCGCATCTCGAGAGTGC..... 11443  
1275 LeuProArg.....LysProSe 1280  
11444 CTCCCTCGAGTCCATCCCGGTGAGCTCGAAGGCTCGACCAACCTCG 11493  
1280 Pro.....ProArgAlaMetAlaSerLeuProG 1290  
11494 GCGACCTCGGCTGCTCGTACGCGCGGAGGCTCGGCTCGCT 11543  
1290 InSerSer.....His 1293  
11544 CTACAGCGCGGCGAGAGCTCTTTCGAAAGCCAGCTCCGAGAGCG 11593  
1294 ProArgArg.....ProProPheLeuProGlnAlaArgLeuSerAlaAlaPhe 1309  
11594 CGGAAAGAGTCCCGCGGCTTCCG.....TCCACCTCGGCTCGAGC 11640  
1309 eGlnGlyAlaMetLysThrLysValLeuProProArgLysGlyArgGly 1326  
11641 AGGCGCTCGATGACTGCGCTTTCGAGAGAGAGAGAGAGAGAGGCTCG 11690  
1326 eu..... 1326  
11691 TCTGCGCTCTCTCGTGAAGAGAGCTCGGCTCGCTCGCTCGACCA 11740  
1327 .....LysLeuGlnAlaIleValGlnLysIleThrSerProSer..... 1339  
11741 CTTGCGCGTGGCTTCACCGCTCCAGAGGATGATCTCTCGTCTCGATC 11790  
1340 .....LeuLysLysPhe 1344  
11791 GTCTGAGCGAGCGCGAGTGAACCTCTTGGCTGCTGAGAGCGCTCGC 11840  
1344 Lacys..... 1345  
11841 CATCGGAGAGAGCTCGCGCGGAGCTCGACCCCGGAGCTTCACAC 11890  
1346 .....LysAlaPro..... 1348  
11891 ATGATGCGCTCTTCGCTGCTGAGTGAAGAGAGCTCCAAACCCGCTTCA 11940  
1348 ..... 1348  
11941 CGGCTGCGCGCGCGAGGCGCTCTTCGCGACAGCGGCGAGATCT 11990  
1349 .....GlyAlaSerProGlyAsnProLeuSer.....ProSerLeu 1360  
11991 CGGCTGAGAGCGCGGCTCGCGGAGAGAGCTCGGCTCGCTCGCAA 12040  
1361 SerAspLysAspArgGlyLeuLys.....GlyAlaGlyLysProValGly 1376  
12041 GCGGCTCGAGAGGCGGCTTCGCTCGCGGCTCGCTCGCTCGCTCG 12085

```

12865 CTCGTCGATCTCGGCATCTGAAACCAATCAACGTGTCCTCATCCATCGC 12914
1634 .....SerSerSerSerSerPheserLeuAspAlaIadgLy 1645
12915 GGCGCGCGGGCGGCTCGGGAGCGGCCCGCTTCAAGTCTGCACGACCACTCG 12964
1646 AlaserLeuAlathLeuPro.....GlyGlySerIleLeuGlnPr 1659
12966 GCGCGGAGGCTTGTGCCACCCGCAAGTCCAGGAGATGGAGGGGTCTCCGC 13014
1659 ArgProSerLeuProLeuSerSerThrMethHisLeuGly...ProValI 1675
13016 GCGCTCG.....GCTTCGAGATGGCACCTCGCGTCTCTCAAG 13052
1675 alSerLysAla...LeuSerThrSerCys 1683
13053 TGACCTGGGCTTCGACGACGACTTCCTCGC 13081
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-764-233A-1
seq_documentation_block:
; Sequence 1, Application US/08764233A
; Patent NO. 5716849
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Schupp, Thomas
; APPLICANT: Beck, James J.
; APPLICANT: Hill, Dwight S.
; APPLICANT: Neff, Snezana
; APPLICANT: Ryals, John A.
; TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-geigy Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,233A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/729,214
; FILING DATE: 09-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: 1506/CTP6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 541-8587
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 49377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Sorangium cellulosum
; IMMEDIATE SOURCE:
; CLONE: p98/1, pJL3, and pVKM15

```

FEATURE:  
 NAME/KEY: misc.feature  
 LOCATION: 383..760  
 OTHER INFORMATION: /product= "SorB"  
 OTHER INFORMATION: /note= "This gene encodes a protein that is highly homologous  
 OTHER INFORMATION: the reductase domains of type I PKs such as eryA from  
 OTHER INFORMATION: Saccharopolyspora erythraea."  
 FEATURE:  
 NAME/KEY: misc.feature  
 LOCATION: 927..19874  
 OTHER INFORMATION: /product= "Sora"  
 OTHER INFORMATION: /note= "gene product is highly homologous to type I PKs that  
 OTHER INFORMATION: are known to be involved in the synthesis of polyketide  
 OTHER INFORMATION: compounds."  
 FEATURE:  
 NAME/KEY: misc.feature  
 LOCATION: 942..7115  
 OTHER INFORMATION: /product= "Module 1 of Sora"  
 FEATURE:  
 NAME/KEY: misc.feature  
 LOCATION: 7203..12884  
 OTHER INFORMATION: /product= "Module 2 of Sora"  
 FEATURE:  
 NAME/KEY: misc.feature  
 LOCATION: 13455..19616  
 OTHER INFORMATION: /product= "Module 3 of Sora"  
 FEATURE:  
 NAME/KEY: misc.feature  
 LOCATION: 19870..24556  
 OTHER INFORMATION: /product= "Module 1 of SorB"  
 FEATURE:  
 NAME/KEY: misc.feature  
 LOCATION: 24638..30820  
 OTHER INFORMATION: /product= "Module 2 of SorB"  
 FEATURE:  
 NAME/KEY: misc.feature  
 LOCATION: 30881..35446  
 OTHER INFORMATION: /product= "Module 3 of SorB"  
 FEATURE:  
 NAME/KEY: misc.feature  
 LOCATION: 35528..40114  
 OTHER INFORMATION: /product= "Module 4 of SorB"  
 FEATURE:  
 NAME/KEY: misc.feature  
 LOCATION: 40190..46318  
 OTHER INFORMATION: /product= "Module 5 of SorB"  
 FEATURE:  
 NAME/KEY: misc.feature  
 LOCATION: 46851..47891  
 OTHER INFORMATION: /product= "SorY"  
 OTHER INFORMATION: /note= "The protein encoded by the sorY gene is highly  
 OTHER INFORMATION: homologous to the methyltransferase from Streptomyces  
 OTHER INFORMATION: hygroscopicus that is involved in the synthesis of the  
 OTHER INFORMATION: polyketide rapamycin."

```
alignment_scores:
  Quality: 493.00      Length: 2095
  Ratio: 0.593         Gaps: 102
Percent Similarity: 39.666  Percent Identity: 22.291
```

alignment\_block:

Align seg 1/1 to: US-08-764-233A-1 from: 1 to: 49377

62 GlyAlaGlyThrProSerGlyThrAlaAlaValAla.....Al 75

[illegible]

24497 CCGGCCACCCCTCGCCTTCGACCATCCCTCTCTCTCAFCAGATCGCGCTCTT 24544  
340 roSerSerHisSerProAlaArgSerValGlyArgSerProSerTyr 356  
::: ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
24547 CTTCGGGACTCGCTGCACCAGCCCTCGCACGAGGCTCTCCGTGS... 24593  
357 SerSerThrProSerProLeuMetProAsnLeuGluAsnPh... ProTy 372  
||| ||| |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
24594 .....AGCCCGACGCGCGCGCTCCGCGCTTCGCGCGCGG 24631  
372 rSerGlnGlnProLeuSerThrGlyAlaPheProAlaGlyIleIthrAspH 389  
AGG:::|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
24632 AGGCGAGCGACATCGCCATCTGGCATGGCCCTCGCCCTCGCGCGGCGG 24681  
389 isSerHisPheMetProLeuLeuAsnProSer...ProThrAspAlaThr 404  
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
24682 CGTCGGCATGTCCAGCCTCTTTGGGAGTTCCTGGCCCCAGGACGCGAG 24731  
405 SerSerValAspThrGlnLagLyAsnCysLysProLeuGlnLys...As 420  
::: ||| |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
24732 GCGTCG.....AGCCATTCCAAGGCGGA 24757  
420 pLysLeuProGluAsnLeuLeuSerAspLeuSerLeuGlnSerLeuThrA 437  
::: ||| |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
24758 TGGGATCGCGCTCGCTTACAGCCCGACCGCCGACGACCAAGAAC 24807  
437 laLeuThrLeuGlnValGluAsnIleSerAsnThrValGlnGlnLeuLeu 453  
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
24808 CTAGCTCGGCATCGCCCATCTCGACAGGTCGACCTCTTCGACCTTG 24857  
454 LeuSerLysAlaAlaValProGlnLysGlyValLysAsnLeuValSe 470  
||| |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
24858 CCTCTTTGGCATCAGCCCGGAGGCGAAAC.....AC 24892  
470 rArgThrProGlnGlnHisLysSerGlnHisCysSerProGluGlySerG 487  
24893 CTCGACCCC.....AGACGCGCTCTCTCC.....TCGA 24921  
487 lyTrSerAlaGluProAlaGlyThrProLeuSerGluPro..... 501  
24922 ATCTGCTGGCAGCGCTCGAGACGCGGCATCTGTCGCCCCACCTCA 24971  
502 .....SerSerThrProGlnSerThrL 509  
24972 AGGATTCGCCACGGGGCTTCGTCGCATCGCGCCAGCAGCAATACGA 25021  
509 salAGlu.....ProGlnGluAlaA 516  
25022 TTGCGAGAGGCGACACCGAAGATTCCGACGCTATTGCCCCCAAGCA. 25070  
516 spTYrLeuSerGlySerGluAspProLeuGluArgSerPheLeuTYrCys 532  
25071 .....CCGCGGGTCTTTTCGGGGGGGCGCTTGCCTACA 25106  
533 AsnGlnAlaArgglySerProAlaArgValAsnSerAsnSerLysAlaly 549  
::: ||| |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
25107 CGCTGCGCTGCAAGGCGCG...CGCTCGGTTCGACACCGCCTGCTCC 25153  
549 sProGluSerValSerThrCysSerValThrSerProAspaspMetSert 566  
::: ||| |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
25154 TCCTCGCTGTCGCGCTCCACCTCGCTCGCAAGCCTTCGACAGGCGCA 25203  
566 hrLysSerAspSerPheGlnSerLeuHisGlySerLeuProLeuAsp 582  
::: ||| |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
25204 GTGCAACCTTGCCCTCGCGCGGGGCTTCGCTCATGGCTTCCCGAGG 25253  
583 SerPheSerLysPheValAla..... 589  
::: |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
25254 GCTTCGTCTCTTTTCCGCGCTTCGGCGCTTCGGCGCCGACGCGCGCTCC 25303  
590 .....GlyIluArgaspCysP 595  
::: |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

25304 ANGACCTTCTCGGCCAACGCGGACGGCTACGAGCGAAGAAGCGCTCAT 25353

595 rOArgLeuLeuSerAlaLeu.....AlaGlnGluAspLeuAla 608  
||| ||||| ::: ||||| |||  
25354 CGTCTTTCCCTCGACGGCTCGGTGCAGCGCCTCCGCCGAGGACACCGCG 25403  
609 SerGluIleLeuGlyLeuGlnAlaIleGlyGluLysAlaAspLysAl 625  
||||| ::: ||||| |||  
25404 TCTCGCCTTCGTCCGGCGCACCGCATCAACACGACGAGCGCGCTCGAGC 25453  
625 a.....TrpAlaGluAlaProS 631  
||| ::: ||||| |||  
25454 GGTATCACCGCCCCAACGGCACCTCCACGACGAGGTCTCTCGGCGCGC 25503  
631 erLeuValLysAspSerSerLysProPheSerLeu..... 643  
::: ||||| ||||| |||  
25504 GCTCCACGACGCCGATCACTCCCGCGGACGCTGACGCTGCTCGAGTGCC 25553  
25593 .....AAGCCCTGCGCG...CGGTCTACGCGCAGCGCAGACCG... 25628  
670 erLeuGlnLeuAspLysGlyGlyAsnAlaLysAspPheSerProGlyLeu 686  
||| ::: ||||| |||  
25629 .....CTGAAGAAGCCTCTCCTCTCGCGC 25652  
687 PheGluAspProSerValAlaPheAlaThrProAspProLysLysThrTh 703  
::: ||||| ::: ||||| |||  
25653 CGTCAAGACCACATCGGCCATCTCGAGCGCGCCTCGCGCTCGCGGCG 25702  
703 rGlyProLeuSerPheGlyThrLysProThrLeuGlyValProAlaProA 720  
::::: ||| ||| ::: ||||| |||  
25703 GTCGCCACAATCG.....TGCCCTCCTCGCCACGACGCGCTCGCGCC 25746  
720 spProThrThrAlaAlaPheAspCysPhePro.....AspThr 732  
||| ::: ||||| |||  
25747 CACCCTCCACAGCGGCGCGCGCAAT....CCCTGATTGATTGGGATACA 25792  
733 ThrAlaAlaSerSerAlaAspSerAlaAsnProPheAlaThrProGluGl 749  
||| ::: ||||| ::: ||||| |||  
25793 CTCGCGCTCAGCTGGTGTGATACC.....CGAGGTTCTGGCGCGCGCA 25836  
749 uasnLeuGlyAspAlaCysProArgTrpGlyLeuHisProGlyLeuLeut 766  
::: ||| ::: ||||| |||  
25837 CGAA.....GATAGAGTCCCGC..... 25855  
766 hrLysGlyLeuGlnGlnGlyLysAlaSerAspGlyLysLysGly 782  
:||| ::: ||||| ::: ||||| |||  
25856 .....CGCGCGCGCTCTCCGCTTGGACTCTCGGCACTCCGCGCAC 25891  
783 AspThrHis.....GluAlaSerAlaCysLeuGlyPheGlnGl 795  
:||| ::: ||||| |||  
25892 AAGCCACAGTATCTCTCGAGAGGCTCCCGCGCGCGCTG.....TC 25932  
795 uGluAspProGlyGlyLysValAlaSerLeuProGlyAspPheLysG 812  
::: ||||| ::: ||||| |||  
25933 GGCGAGCGGCCACCTCAACAGCGGCTCGGACGCTCCCGCGCGCT 25982  
812 lncGlnValGlyGlyValLysGlnGluAlaGlyLysLeuGlnCys 828  
::: ||||| ::: ||||| |||  
25983 GTCCGCTGCTCTGTGCGCCAGAGGAGCGCGCTCGCGCGCCAGCGG 26032  
829 ProGlu.....ValAlaAlaAlaAspArgTrpLeuGluAs 840  
::: ||| ::: ||||| |||  
26033 AAGCGGCTCCGACCACTCTCGCCACGACGACCTCGCCTTATCGA 26082  
840 pSerArgHisCysCysSerThr.....AlaAspPhe.GlyAspLeuPro 854  
::: ||||| ::: ||||| |||  
26083 TGTTGCCATTTCGAGCGACCAACCCCGCGCCACTTTCGAGCAGCGCGCG 26132

26922	CCGACACAGAGCTGACAGGCTCCAGCTCCAGCCGAGCGGCTACTGGATCGA	26971
1114	.....AlaLeuProValAlaSerAspSerProMetCysLeuTyr	1128
26972	AACTCCGCGCAAAACCGCTCTGTCTTCAGAGCGACGACGAGCGCTCCGCA	27021
1128	hLeuSLeuThrAspSerProSer	1135
27022	CGATTGGCATCGCTCTTCGCTGAGAGTACGACCCCATCCGCTGACGC	27071
1136	ThrProGlyLysAspAlaArgSerMetIleLeuArgSerAlaThrLeuTyr	1152
27072	TCCGCTCCGCGGACACCTGAGAGGCTCCACGCGTACGATCCGCTGCTGC	27121
1152	hLeuSLeuLeuPheHisSerTyrAlaArgArgProSerLeuGlyAlaLeu	1168
27122	GG.....CTCCATTTCAGAGAAAGAAAGGCACTCG.....	27152
1169	ProAsnCys.....	ArgAla
		1173
27153	CCGCGCTCTCTCTCTCTCTGCGGCGAGGCTCTTCACCCGAGGCTCGGCTC	27202
1173	AlaThrLysLeuLeuAspAsnSerHisLeuProAlaThrPheLeuVal	1190
27203	GACTGGAAGACTCTTCGCGGCGGCAAGCTCCCGGGAAGTGTCCCTCC	27252
1190	ArgSerSerProGlnLysGln.....GlyArgValSerGlnArgAla	1203
27253	CAACTACCCCTTCAGAGAGAGAGCGGTCTGCTCGCTCGACGTCCAGAGAGC	27302
1204	ArgAla.....	1205
27303	AAGCTTCGAGACGTGCTCCGAGGCTGACTCGGCGGACCAATCCG	27352
1206	.....ProLysProGlyAla.....	1210
27353	CTGCTGGGCGCGCGGTGCGCTTCGCGACGCGGCGGTCTCTCTTAC	27402
1211	.....GlySerLysLeuSerAspArg	1217
27403	AGGCGGACTCTCCCTCGACAGACACCGGAGCTCGAAGGCAATGCGCT	27452
1218	ProLeuHisAlaLeuLysArgLysSerAlaPhe.....	1228
27453	TGCGCACACCATCTACCGGAGCACCGGCTCTTCGAGCTCCGCTCGAC	27502
1229	.....MetAlaProValProThrLys.....LysArg	1238
27503	GTCGCGCACCGGCGTGGCTCGACACCGTCGAGAAAGATCAACGCTGAGAGC	27552
1238	hLeuValLeuThrGlnSLeuSerSerSerSerSerAsnAlaSerAlaMet	1254
27553	CCGCTCGCTCCCATCGAGACACGCTGCTCCAGACTCCGCTCG	27602
1255	Gly.....	1255
27603	GACCCCTGAGACGACGAGAGGAGGCGCTCTTCATACGCGGACA	27652
1256	.....GluMet	1258
27653	GAGACCGCGTTTCAGAGTGGCCCTGACTGGACACCGACGGGCTCT	27702
1258	LysArgArgArgGlyLeuArgValProProProSerSerArgGlyCysLeu	1274
27703	CTGCGCGGAGCAACCATCCCTCTCCGCGCATCTCCAGAGTGGC.....	27744
1275	LeuProArg.....LysProSe	1280
27747	CTCCCTCGATGCGCATCCCGGTGAGACTGAAAGCGCTTCAGACACCTC	27796
1280	rPro.....ProArgAlaMetAlaSerLeuProG	1290

27797 GCCAACCTGGGCTTTCCTACAGCCCGAGTTCACAGGCTCCGCTCCGT 27846  
1290 InSerSer.....His 1293  
: : : : :  
27847 CTACAGCGCGGAGAGCTCTTCGCGAGCCACAGCTCCCGGAGCGG 27896  
1294 ProArgArg...ProProPheLeuProGlnAlaArgLeuSerAlaAlaph 1309  
: : : : :  
27897 CCGAAAGGATCGCGCGCTTTCGCC...TCCACCTCGCTCGTCCGAC 27943  
1309 eGlnGlyAlaMetLysThrLysValLeuProProArgLysGlyArgGlyL 1326  
: : : : :  
27944 AGCCCTCGATGACCTGGCTTGGAGCAGCAGAGAGGAGCTCG 27993  
1326 eu..... 1326  
: : : : :  
27994 TCTGCCCTTCTCGTGGAGCGAGTCTCGCTCGCTCGTGGTCCACCA 28043  
1327 .....LysLeuGluAlaIleValGlnLysIlePheSerProSer..... 1339  
28044 CTTGGCGGTGGCTTCCACGCTCCAGGGGTGATCTCCCTCTCGATC 28093  
1340 .....LeuLysLysPheA 1344  
: : : : :  
28094 GTCTGCGCGAGCGCGAGCTGACCTCTTCTCGCTCGTSCAAGCGCTCG 28143  
1344 lacys..... 1345  
: : : : :  
28144 CATCGGAGCAGCTCGCGCGAGCTCCGACCGCGGAGCTTCCAC 28193  
1346 .....LysAlapro..... 1348  
28194 ATGATGGCTCTTCGGCTGCTGAGCGAGCTCCAAAGCGCCACTCA 28243  
1348 ..... 1348  
28244 CCGCTGCGCGCGCGAGCGGCTCTTCGCGACAGCGGCGCAGATCT 28293  
1349 .....GlyAlaSerProGlyAsnProLeuSer...ProSerLeu 1360  
28294 CGGCTCGAGCGCGCTCGCGCTGAGCGGAGCTGCTGCTCCCTCCGAA 28343  
1361 SerAspLysAspArgLysLeuLys...GlyAlaGlyLysProValG 1376  
: : : : :  
28344 GCGGCTCGACCGAGGCGCTTCGCTCGCGGCTCGCTCGCGCC... 28388  
1376 yValGlnGlyLeuValAsnValGlyThrGlyGlnLysLeuProThr. 1392  
: : : : :  
28389 ...CCTCATGATCGACCGCGAGCGAGCTCTCCGAGCGCCACGA 28434  
1393 .....SerGlyAlaAspProLysArg.....AsnProThrAsn 1404  
28435 GGCACGCGCTCGACCTCGCTCTTGAAGCTGCTGCGCGAGCAGAC 28484  
1405 ArgSerLeuLysGlyLysLeuMetAsnSerLysLysLeuSerSerThrAs 1421  
: : : : :  
28485 GCCTCGCTCGCGGCTGCTGCTCGTCAACCGAGCGCTCGCCACC 28534  
1421 pCysPheLysThrGlnAlaPheThrSerProGluAlaLeuGlnProGlyG 1438  
: : : : :  
28535 CACACGAGAGCAGCTCAGAGCTCG.....CTCAGCGCGCGCTG 28578  
1438 lyThrAlaLeuAlaProLysArgSerArgLysGlyArgAlaGlyAla 1454  
: : : : :  
28579 GGGCTCGCGCTCGCGGAAAGTAGACACCGACCTCCCGCTCTCC 28628  
1455 HisGlyLeuSerLysGlyProLeuGlnLysArgProTyLeuGlyProAl 1471  
: : : : :  
28629 TCGTCGACATCG.....ACCTCAGCGAGGCTCCACGAGCGCTG 28669  
1471 alaLeuLeuThrProArgAspArgAlaSerGlyThrGlnGlyAlaSerG 1488  
: : : : :  
28670 CTAGGCGCGCTCGACAGAGAGACCGCAGCTCCCGCTCCGCAAGGAA 28719

1488 luAspAsnSerGlyGly..... 1493  
: : : : :  
28720 ACCCTCATCCGAGGTGGCGACACCGCTCGACGAGCGCTCATCC 28769  
1494 .....GlyGlyLys.....LysProLysMetGluG 1502  
: : : : :  
28770 CGCGCAAGCACCCAGCTGGCGCTCCATATTCGACCAAGGACCTTC 28819  
1502 uLeu...GlyProAlaSerGlnProGluGlyArgProCysGlnProG 1518  
: : : : :  
28820 GAGCGCTCGCTCGTGGAGCGCCCG.....AGGCCAGGCGCCCT 28863  
1518 lnThrArgAlaGlnLysGlnProGlyHisThrAsnTySerSerTySer 1534  
: : : : :  
28864 CGCACAGGCCAAGTCCGCTCGCTGACCGGCGGAGGCTCAACTCC 28913  
1535 LysArgLys.....ArgLeuThrArgGlyArgAl 1544  
28914 GCGATGCTGCTCGACACCTTGGCATGTATCGGGCGACGCGCCGCG... 28961  
1544 aLysAsnThrThrSerSerProCysLysGlyArgAlaLysArgArgG 1561  
: : : : :  
28962 .....TCGGAGCGGAGCGCGGCTGTTACTGAGT 28995  
1561 lnGlnGlnValLeuProLeuAspPro..... 1569  
: : : : :  
28996 CGGTCCAGGTGCTCCGATACACGCTAGCGACCGGCTGATGGGGTCT 29045  
1570 ...AlaGluProGluIleArgLeuLysTyIleSerSerCysLysArgLe 1585  
: : : : :  
29046 TCGCGCGACCTTGGTC..... 29063  
1585 uArgSerAspSerArgThrProAlaPheSerProPheValArgValGluL 1602  
: : : : :  
29064 .CCACGCGCATCGCGAGCGCGCGCTGATCT..... 29093  
1602 ysArgAspAlaPheThrThrIleCysThrValValAsnSerProGlyAsp 1618  
29094 .....GCCCATCCCCACGCTGCTCT 29117  
1619 AlaProLysProHisArgLysProSer..... 1628  
29118 TCGCCCAAGCGCGAGCTGCTCCCATCACTATCTACCGCTTACTATGGA 29167  
1629 .....SerAlaSerSer..... 1633  
29168 CTCGCTGATCTCGGCTATCTGAACCAATCAAGTCTCTCATCCATGC 29217  
1634 .....SerSerSerSerPheSerLeuAspAlaAlaGly 1645  
29218 GCGCGCGCGCGCTCGGAGCGCGCTTCACTGCTGACGCGCACCGCTCG 29267  
1646 AlaSerLeuAlaThrLeuPro.....GlyGlySerIleLeuGlnPr 1659  
: : : : :  
29268 GCGCGAGCTCTTTCGACCGCGCTGACGAGGAAAGTGGAGCGCTCTCGC 29317  
1659 oArgProSerLeuProLeuSerSerThrMetHisLeuGly...ProVal 1675  
: : : : :  
29318 GCGCTCG.....CTTCGAGCATGCGACCTCGCTCTCTCAG 29355  
1675 alSerLysAla...LeuSerThrSerCys 1683  
29356 TGACCTGGCTTCGAGCAGCAGCTCTCTCG 29384

seq\_name: /cgn2\_6/ptodata/2/lna/6B\_COMB.seq:US-09-144-085-3

seq\_documentation\_block:  
; Sequence 3, Application US/09144085  
; Patent No. 6280999  
; GENERAL INFORMATION:  
; APPLICANT: Gustafsson, Claes  
; APPLICANT: Betlach, Mary C.



```

13009 AGGCTCGGCGCT.....GGCTCGGCTCCCGGCAACACACC 12969
477 .....LysSerGlnHisCysSerProGlu.GlySerGlyTyrSerAl 490
12968 CGGCGTAAAGCTCGGCTCGACGCCAGAGCTCCAG..... 12930
490 aGluProAlaGlyThr.....ProLeuSerGluPro..... 501
12929 .AGGCCCTGAACACCGAGCGTAGTCGAGCCGCGCCCTCAGCAGCCGCT 12881
502 .....SerSerThrProGlnSerThrHisAla 510
12880 GGTAAACCCGCGAGTCCACCTCTCTCGCGCCGCGCAGACCCACGTC 12831
511 GluProGlnGlu.....AlaAspTyrLeuSerGlySerGluAs 523
12830 GACAGCTCGTCGAGCTCGCGGGGCGCGGGATCTCGTCCGTCACAC 12781
523 pProLeuGlu.....ArgSerP 529
12780 CCGCGTGGGCTGTCACCCACGCGCGCTCTCGGCGCCTGCTCGGACT 12731
529 heLeuTyrCysAsnGlnAlaArg.....GlySerProAla 540
12730 GCGTGTACAGCTTACCTTCGCGCGCGCGCGCGCTCGGCGCTCCGACC 12681
541 ArgValAsnSerAsnSerLysAlaLysProGluSerValSerThrCysSe 557
12680 GACAGCTGTCAGCGCGCGCGCGCTCTCTCGGCGCAGCACACGCGCTCGGC 12631
557 rValThrSerProAspMetSerThrLysSerAspSerPheGlnS 574
12630 CAGAACCGAGCTCGACAGCGTCCGCGTCCGCGCGCGCGCGCGCGCGCA 12581
574 exLeuHisGlySerLeuProLeuAspSerPheSerLysPheValAlaGly 590
12580 GCGCCAGCTCCAGCATCCCG.....GTGCCGGA 12552
591 GluArgAspCysProArgLeuLeuLeuSerAlaLeuAlaGlnGluAspLe 607
12551 AACACCCACTTGCACCAACACCTGATGATCCCGA.....AGCCACGATGCTC 12505
607 uAlaSerGluLeuLeuGlyLeuGlnGluAlaIleGlyGluLysAla.... 622
12504 CCGCAGCAGCAGCTCCCTGTGTACAGAGTGCCCTTCGCGCTCGGCGAGT 12455
623 .....AspLysAlaTrpAlaGluAlaProSerLeuValLys 634
12454 TCGTCGCGCGCGCGAGCAGCGAGTGAGCGCGCTTCAGCGCGCGCGAG 12405
635 AspSerSer..... 637
12404 CCGAGTTCGCGCGCGCGCTCGGCGCTCCAGCCAGTAGCAGCTCCGCGTG 12355
638 .....LysProPheSerLeuGluA 645
12354 GAACCGTAGTCGCGCAGCTCCAGCGCGCGCGCGCTTCAGCTGACCGCTCGC 12305
645 snHisSerAlaCysLeuAspSerValAlaLysSerAlaTrpProArgPro 661
12304 CTTGGGCGCAGTCCACTCGTGCCTTCAGCTGACCGCTCGCAGCGCGG 12255
662 GlyGluProGluAlaLeuProAspSerLeuGlnLeuAspLysGlyLys 678
12254 TCACAAGCGCGCTCTTCTCTGCTCTC.....GCCGAG 12217
678 nAlaLysAspPheSerProGlyLeuPheGluAspProSerValAlaPheA 695
12216 CTCGCCCAACACGTCGCGCTCGGCTC.....CCTCGGCA.....G 12182
695 laThrProAspProLysLysThrThr..... 703

```

```

12181 GCACCCCGCCCGCAGCGCGCACAGCAGCCATCCGAGCGCACTCGACGT 12132
703 ..... 703
12131 ATGTGCTCACCCCGCGCGCGCGCAAGCGTGGGCATCCCTCCAGGAAGCGC 12082
704 .....GlyProLeuSerPheGlyThrLysProThrLeuGlyVal. 716
12081 ACCGCTCGGCGCACTTCGCTCACCAGTACTCGGCGGACATCAGCGATT 12032
716 ..... 716
12031 TTCGCCACCGAGCTCGCGCTCACCGCGCTCACCGGGTACCTCGCGGC 11982
717 .....ProAlaProAspProThrThr.....AlaAlaPh 726
11981 TCCCATAGCGCACTCCCGCGCACCTCCCGCTACTCTTCCAGCATCCCG 11932
726 eAspCysPheProAspThrThrAlaAlaSerSerAlaAspSerAlaAsn 743
11931 TCCATGTCGCGGCTGTGGAATCGTGGACACCGCGGACGCGCG..... 11886
743 rPheAlaTrpProGluGlnAsnLeuGlyAspAlaCysProArgTrpGly 759
11885 ..TGCGCGCGCGCTTCGCGCTCGAACCGCTGCCACCTCCAGCACCGCGC 11838
760 LeuHisProGlyGlnLeuThrLysGlyLeuGlnGlyGlyLysAlaSe 776
11837 CTTGCTCCCGCTCAGCACCGCTCTGCTCGGACCGTGTGATCGCGCGATC 11788
776 rAspGlyIleSerLysGlyAspThrHisGluAlaSerAlaCysLeuGlyP 793
11787 CCAGCGCGCGCTCCAGCCCGCTCCCGCAGCG..... 11754
793 heGlnGluAspProGlyGluLysValAlaSerLeuProGlyAsp 809
11753 .....GCCCTGCACCTCGGCTCGGCTCCAGCGACA 11718
810 PheLysGlnGluGluValGlyGlyValLysGluAlaGlyGlyLeuLe 826
11717 TCATGCTCCCGCGCGCTGGC..... 11697
826 uGlnCysProGluValAlaLysAlaAspArgTrpLeuGluAspSerArgH 843
11696 .....ACCTCTGCATCAGCGCGCGC.....GGCGCA 11669
843 iCysCysSerThrAlaAspPheGlyAspLeuProLeuProThr 859
11668 CACTAGCTTCGCTCGCTCGCGCAAGGCTCA.....GCAGCGCGCGCAGT 11625
860 SerArgLysGluAspLeuGluAlaGluGluLysThrSerSerLeuCysG 876
11624 GCGCAGCGCTCA.....GC 11611
876 uLeuLeuGlySer.ProGluGlnArgProGlyMetGlnAspProLeuSer 892
11610 TCTCTATCGAGTCCCGCAGCAGCGCGCGG.....CCTCAGTCC 11570
893 ProLysAlaProLeuIleCysThrLysGlu..... 902
11569 CCAGCGTCCCGCTCGCGGTACAGCGCGCTTCAGCGCAACACCGCG 11520
903 .....GluValGluGluValLeuAspSerLys.....AlaGlyTrp 915
11519 GCTGCTGTACTTCGCTCGCTCCAGCGCGCTCTCTGCTCGCTCGCGCT 11470
915 lYSerProCysHisLysSerGlyGluSerValIleLeuLeuGlyProThr 931
11469 TCCGCGACACCACTCTCT.....CAACCCAG 11441
932 ValGlyThrGluSerLysValGlnSer..... 940
11440 GTC.....GAGATGCGCTCCAGCGCTTCGACACCTCGTGGAGCGCG 11397

```



941 .....TrrpNegUserSerLeuSerHisMetLysProG 952  
11396 CAGGGAACAGGAGGAGTACTGTGTAAGTCTCTCCACATCCGACAGCCG 11347  
952 LysLugLugLysProAspGlyLysArgAlaProGlyAspSerThrThrSer 968  
11346 TGGCTCCCTCCCTCCCTGTAACAGCAC..... 11321  
969 AspAlaSerLeuAlaGlnLysProAsnLysProAla.....ValProG 983  
11320 ...CGCAAGCTTCCCCCGCGGCTCCGCGGACGACAGCACTGCGG 11274  
983 uAlaProLleAlaLysLysGluProValProArgLysSerLeuArgS 1000  
11273 CGGTCGACCCCTCCGACAGCGGCTCAGCCCTCCGACGCGCTCCGACAG 11224  
1000 eArgArgValHisArgGlyLeuProGlnAlaGlnAspSerPro CysArg 1016  
11223 CTGCGCGCTGCAACGATCGCGGCA...GGCAAGTCCGTCCGCTGCA 11177  
1016 gAlaProValLeu.ProLysAspLeuLeuProGlnSerCysThrGly 1032  
11176 CGCGCGCTCTCCACCACT.....CGGCGACCGCA 11145  
1033 ProProGlnLysLysMetLysLysAlaProGlyArgGlyAlaSe 1049  
11144 CCTCCGCGCTGCT...CTCCGACCACTTCCGACGCGCTCCGCTGCGCC 11098  
1049 rGlnLysLeuProArgMetCysThrArgSerLeuThrAlaLeuSerGlnP 1066  
11097 CCCACCGAGGCTCTGCTGCGCCG.....ACAGCAACAGCGGAGT 11057  
1066 rArgThrProGlyProProGlyLeuThrThrPro.....Ala 1079  
11056 CGGACACCTCCGCGCGCAAGCGTCCGACGCGCGCGCTGCGCA 11007  
1080 ProProAspLysLeu..... 1084  
11006 CGGCTCAAGCGCGCTCTTCCAGAGTACAGCGCGCTCCGCGCTG 10957  
1085 .....GlyLysLysGlnArgAla 1090  
10956 ATGCCAAGACAGCAACAGCGCGCAAGGAGGAGCGCTGGGCGCGCG 10907  
1091 ..AlaPheLysSerGlyLysArgValGlyLysPro.....SerProLys 1104  
10906 AGGCGACGCGGTACGAGGCTGACACCTGACCGCGCAGACTCTCCAGG 10857  
1105 .....AlaAlaSerSerProSerAsnProAl 1113  
10856 AGATCTGGGGGTGGCGGCGAGCTGTGCAACGTCGCGGAGGGGCTCA 10807  
1113 aAlaLeuProValAlaSerAspSerProMetGlySerLysLysG 1130  
10806 TGGC.....GAACGCGCGCAGAGATCTTGACACACTGC 10775  
1130 LuThrAspSerProSerThrProGlyLysAspGlnArgSerMetLleLeu 1146  
10774 GGGGATGCGCGCGCGCGCTGACTGAGTACCGAGTGTCTTGCACGCTC 10725  
1147 ArgSerArgThrLysThrGlnLysLlePheHisSerLysArgAlaArgP 1163  
10724 CAAGCC..... 10719  
1163 oSerGlnLysArgLeuProAsnCysArgAla.....ThrLysLys 1177  
10718 .GCAGCGGCGAGATCGCTCCCTGCTGCGGTAACAGCGCGGAGCGC 10670  
1177 eLLeuAspAsnSerHisLeuProAlaThr..... 1186  
10669 TTGCACTCGATCGATGCGCCAGAGAGCGTCCCGTACCGTGGGACTCGA 10620

1187 ....PheLysValSerSerSerProGlnLysGlnArgValSerGlnP 1202  
10619 CCAAGTGGATGCTGAGAGGCTTCA.....GCCCGCG 10588  
1202 gAlaArgValProLys.....ProGlyAlaGly..... 1211  
10587 TTCCGACAGCGCGCGGACACACCTTCTGCTGGGCGCTGCGCTTGGCGCG 10538  
1212 ....SerLysLeuSerAspArg.....ProLysHisAlaLeu 1222  
10537 CGTGAATCCCTGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10488  
1223 LysArgLysSerAlaPheMetAlaProValProThrLysLysArgAsnLe 1239  
10487 CGCCAGAGACCGGATCCCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 10449  
1239 uValLeuArgHisGlySerSerSerSerSerAlaSerAlaMetLys 1256  
10448 .....GAGCGATGACGACACAC 10430  
1256 LuMetGlyArgArgArgGlyLeuArgValProProProSerSerArgGly 1272  
10423 CCCGAGCGCTCCGCGCGGCGCTGACGCGTGGCGCGCTGCGAGAACGCTG 10380  
1273 CysLeuLeuProArgLysProSerProProAlaMetAlaSerLeuP 1289  
10379 TGCACCGTCCGCTCG.....CGCAAGCCGACGCTGCGGACAGCAGC 10336  
1289 oGlnSer.....SerHisProArgArgProProPheLeuP 1301  
10335 ACMAACCGCGCGGCTTCCGAGACACTGACAGCGCGCGCGGACAGCAGC 10286  
1301 rGlnAlaArgLeuSer.....AlaAlaPheGln 1310  
10285 CTGCAATCGCGCGCGCTGACAGCGTGCAGAGGAGGAGGAGGAGGAGGAG 10236  
1311 GlyAlaMetLysThrLysValLeuProProArgLysGlyLysGlyLeuLys 1327  
10235 CGAGAGCTGCAGCGCGCTATC.....GAGCGACACCGGAGGCGCTTGC 10192  
1327 sLeuGlnAlaLysValGln..... 1333  
10191 GCCCGAGTGAATAGCCACAGCGCGCGCTGCGAGTCCGAGTACGTCGAGA 10142  
1334 .....LysLleThrSerProSerLeuLysLysPheAla.CysLys 1346  
10141 GTCAGCGATAGCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10092  
1346 sAlaProGlyAlaSerProGlyAsn.....ProLeuSerPro..... 1358  
10091 CTGCTGCGCGCGCGCGCGGACAGACACCGCTGCTGAGCCCTTCCAGCT 10042  
1358 ..... 1358  
10041 GCTGTGAGACACTCCCGGCTCTCCAGCGCTCCAGCGAGTCTCCAGC 9992  
1359 .....SerLeuSerAspLysAspArgGlyLeu..... 1367  
9991 AGCAGCGATGCTCGGATCCATCCGCTCCGCTCCGCGGAGCTTATCC 9942  
1367 ..... 1367  
9941 GAAGAACCGCGCTCAAGAGATGCCAGCTGCTCCAGACAGGAGGAGCT 9892  
1368 .....LysGlyAlaGlyLysSerProValGly 1376  
9891 TCACGAGCTTTCCTTGGCTTGGGCTTGAAGATCGATATGCTTCC 9842  
1377 ValGlnLysLysLeuValAsnValGly..... 1385  
9841 ACCTGAAACCGTGGTCCGATGCGCTTACGATCCCGCCCTCCAC 9792  
1386 ....ThrGlyLysLeuProThrSerGlyAlaAspProLeuLysArgA 1401

```
9791 GAGCATTGCGCAGAGCTCTCCAGATGTTCTCGCTCCGCGCGCATGGCA 9742
|||||
1401 snProThrasArgSerLeuLysGlyLysLeuMetAsnSerLysLysLeu 1417
|||||
9741 GCCCAGGCCCGCATCGATCGATCGCGCGCTCCGCTCCGGTCCAGGCTC 9692
|||||
1418 Ser.....SerThrAspCysPheLys 1424
|||
9691 AGCGCGTCTCGTCCGCGCGCTCGCTCCGCGCGCATGCTCCAG 9642
|||||
1424 sThrGluAlaPheThrSerPro.GluAlaLeuGlnProGlyLysLeu 1440
|||||
9641 CAGCGCGCGCTTACCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 9592
|||||
1441 LeuAlaProLysLysArgSerArgLysGlyAla..... 1452
|||||
9591 TTGTACCGACACCCCGCTCCGCTCGCTCGCTCGCTCGCTCGCTCGCT 9542
|||||
1453 .....GlyAlaHisGlyLeuSerLysGlyProL 1462
|||||
9541 GCATCAGCAGTCCAGCCCGAGTCCAGAAAGCTCGCTCCGGTCCAG 9492
|||||
1462 euGluLysArgProTyrLeu...GlyProAlaLeuLeuLeuThrProArg 1477
|||||
9491 CCGCTCGGATCTTACGCCAGCAGCAGCGCGCTCGCTCGCTCGCTCG 9442
|||||
1478 AspArgAla.....SerGlyThrGlnGlyAla.SerGluasp..... 1489
|||||
9441 GCTCGCGCAGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 9392
|||||
1490 .....AnSerGlyGlyGlyLysLysPro 1498
|||||
9391 TCCCGCAGCGCTTCTCCCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 9342
|||||
1499 LysMetGluGlu.....LeuGlyProLysSerGlnProProGln 1511
|||||
9341 GCCTTCCAGCGATCGCGCTTGTCTCGATCCGTCAGCAGCGGCT... 9294
|||||
1511 uGlyArgProCysGlnProGlnThrArgAlaGlnLysGlnProGlyHis 1528
|||||
9293 ...CGCGCGCGCGCGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCG 9248
|||||
1528 hAsnTyrSerTyrSerLysArgLysLeuThrArgGlyAla 1544
|||||
9247 TC.....CGCGATCCCGCAGCGCTCG 9225
|||||
1545 LysAsnThrThrSerProCysLysGlyArgAlaLysArgArgGln 1561
|||||
9224 CCGCAGCCCACTCGAGCGCGCAAGCGCTTGTCCGGGAGATCGGCA 9175
|||||
1561 nGlnGlnValLeuProLeuAspProAlaGlnProGluLysLeu..... 1576
|||||
9174 CCAG.....CCCGCGCTCGCGAGCTCGCGCTCGGCTCGGCTCGG 9143
|||||
1576 ..... 1576
|||||
9142 TCGTGGTCCAGATCCCGCTCCGCGCAGCGCGCGCGCGCGCGCGCG 9093
|||||
1577 LysTyrLysSerSer.CysLysArgLeuArgSerAspSerArgThrProA 1593
|||||
9092 CGCGCTCGCTCGCGCAGCGCGAGCGCGCGCGCGCGCGCGCGCGCTG 9043
|||||
1593 laPhe...SerProPheValArgValGluLysArgAspAlaPheThrThr 1608
|||||
9042 CGTTCGCGCGCGGTAC..... 9026
|||||
1609 IleCysThrValValAsnSerProGlyAspAlaProLysProHisArgLy 1625
|||||
9025 .....CCGCGCTCGCTCGCGCGCGCGCGCGCGCGCGCGCGCG 8997
|||||
1625 sPro.....SerSerAlaSerSerSerSerSerSerSerS 1638
|||||
```

```
8996 AGCGATGCTGCGCAATAAGCAGCAGCAAGCATCCAGCTCGCGCTGCGCAGCA 8947
|||||
1638 er.....PheSerLeuAspAla..... 1643
|||||
8946 GCTCTTGAGAGTGCCATGCTCGCTTGACCTTCGCGGAGCTCTCTGCGCC 8897
|||||
1644 .....AlaglyAlaSer.LeuAlaThrLeuProGlyLysSerLysLeuG 1658
|||||
8896 AGCATCTCGCGCGAGCTCTCGCACCGCGCAGCGCGGA.....CTAT 8853
|||||
1658 lnProArgProSerLeuProLeuSerSerThrMet..... 1669
|||||
8852 CCGCGCGCATCGCGCACCGCTCAGCGGCTCTGCTCTCTCTCTCTCTCTCT 8803
|||||
1670 .....HisLeuGlyProVa 1674
|||||
8802 CCCTCAGCAGCGCTCGACCTGCGCTCGCATCATCGCAGCGCGCA 8753
|||||
1674 lValSerLysAlaLeuSerThrSerCysLeuValCysLysLeuGlnA 1691
|||||
8752 AGCGTCACGCGG.....C 8739
|||||
1691 snProAlaAsnPheLysAspLeuGlyLysLeuCysGlyProTyrTyrPro 1707
|||||
8738 GCCCGCGC.....CCTCAGCTCTCT 8719
|||||
1708 GluHisCys.....LeuProLysLys..... 1714
|||||
8718 CCGCAGCTCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8669
|||||
1715 .....LysProLysLeuLysGlyLysValArgProG 1725
|||||
8668 ACCAGGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8628
|||||
1725 LuGlyThr.....CysGluGlu 1730
|||||
8627 AGCGTCCGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8578
|||||
1731 AlaSerLeuProLeuGluThrLeuLysGlyPro...GluCysAlaAl 1746
|||||
8577 GCTTCCACCGCGCTCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8528
|||||
1746 a.....AlaAlaThrAlaGlyLysProPro 1754
|||||
8527 ACAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8497
|||||
seq_name: /cgm2_6/ptodata/2/ina/6A_COMB.seq:US-09-335-409-1
seq_documentation_block:
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-335-409-1
alignment_scores:
Quality: 468.00 Length: 2079
```

Ratio: 0.565 Gaps: 113  
Percent Similarity: 39.827 Percent Identity: 22.799

## alignment block:

US-09-508-821b-6 x US-09-335-409-1/rev ..

Align seg 1/7 to reverse of: US-09-335-409-1 from: 1 to: 68750

```

26 ArgLeuGlnAsnTyrArgGlnProSerGlnAlaGlyLeuSerGysAsp 42
|||||
28171 CGGCTACGCCGCCACGACCGCCACACGCG.....CG 28140
42 GlnArgLeuLeuAlaLysAspTyrTyrAsnProGlnProTyrProSer 59
|||||
28139 TCAGCGCGTACTCCACCGTGA..... 28118
59 YrGlnGlyGlyAlaGly..ThrProSerGlyThrAlaAlaValAla 75
|||||
28117 ..GAGCGCGGCGTGGTGAACGCGCTGCTGTCAGACACACGACT.... 28074
75 LaAspLysTyrHisArgGlySerLysAlaLeuProThrGlnGlnGlyLeu 91
|||||
28073 .....CGCGCTCCCGGCTCCGCCACATCCACTCGCGCA 28038
92 GlnLysArgProAlaPheProGlyTyr..... 100
28037 GCGGCGCGTCCAGT...CCGCTCGAACAGCGCCACGACCGGTCGAAC 27991
101 .....GlyValGlnAsp 105
27990 GCCTCCGGAACGCTGGCCACGCGCCGCAAGCCCGCCGATGCCCG 27941
105 eSerProTyrProGlnArgTyrAlaGlyGlnLysLeuAla... 120
27940 CATTGCGCGCGCTGTCGGTGAACAGAACGCCACTGCCGCGCACG 27891
121 .....TrpGlyAlaProGlnPro.....ProProGln 130
27890 AGCTCGGATGACAGCGCGCGCCCGCGCGCTGCCCTGCGCACG 27841
130 nProGlnProLeuProAlaGlyValAlaLysTyrAspGlnAsnLeu 147
|||||
27840 GCCGAAAGCGCGCCACGACG.....CCT 27815
147 yLysThrAlaValProPro.....SerArgGlnTyrAla 158
|||||
27814 GCGCAGCTCACGCCACCGCAGCGCGTCAATCGCGCTGCGCTG 27765
159 GlnGlnGly.....AlaGlnValProPheArgTyr 169
|||||
27764 TCGCAGGCTGAAGCGCAGCTCCCGAGCGCGATCGCATGTC 27715
169 sSerLeuHisValGlnGlnProPro.....ProProGlnGln 181
|||||
27714 AGTGTCTCGCCACCGCGCGCTGCGCATCGAGCGCGTCTGCTG 27665
182 .....ProLeuAlaTyrProLysLeuGlnArgGlnLysLeuGln 194
|||||
27664 GCCCGACAGCACAAAGCTCCCGGAGCGCTCGCGCGGAGGACCA 27615
195 AsnAspIleAlaSerProLeuProPheProGlnGlyThrHisPhePro 211
|||||
27614 GCTCCACCGCGCGCGCTCTCCACACCA..... 27585
211 nHisSerGlnSerPheProThrSerSerThrTyrSerSerValGln 228
|||||
27584 .CATGCGGTTCGTCGCTCATCCGACAGAGCTCAACCCCGCGAGCG 27536
228 LysGlyGlnGlnGlyAlaHisSerTyrLysSerThrAlaProThrAla 244
|||||
27535 CGGCGGTCTCGTGCAG.....GCCAGCGACCGCTGCGC 27501
245 GlnProHisAsp.....Ar 249

```

```

27500 GGGCCACCGCCACCGGACGCGCTCCACGACGACGCGGTAGCTCG 27451
|||||
249 gProLeuThrAlaSerSerSerLeuAlaProGly GlnArgValGlnAsn 265
|||||
27450 CCGAGCTCGCGCTGGCTGGATATTTCCTTGGCCCGACGACGACGCG 27401
266 LeuHisAlaTyrGlnSerGlyArgLeuSerTyrAspGlnGlnGlnGln 282
|||||
27400 CTTCGAG.....CAAGCGCGCCAGCGCGCGCGCTCCATGTCGCA 27357
282 n.....GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 294
27356 GATTGGCGCTGGCGGCTCCAGATACGCGCTGCTCCGACGCGCGCT 27307
294 LysArgHisHisAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 310
27306 TGGCGGTACGCTGCTGACGAGG.....CACTCGATGGGTGCGCCAG 27260
311 Tyr..... 311
27259 CGCGCTCCCGGCTCCCGTGCATCCAGAAATGACGTGCGCGAACA 27210
312 .....GlnHisTyrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 319
27209 CCGCTCGCGCGCCACGCGCTCGCGTGAACGCGCTCTGCGCAGGCGC 27160
319 LysGlyTyrGlnProAspAlaAlaVal.....ArgThr 330
27159 CTGGGACGTGACCGCGCTGCTGCGCGCGCGCGATGATGATGCGCTAC 27110
331 ProGln...GlnTyr..TyrGlnThrPheSerProSerSerHisSer 346
|||||
27109 CCGGATCACCGCGACGAGATGAGGCTCGCGCTGCGCTGCGTGGAGAC 27060
346 roAlaArgSerValGlyArgSerProSerTyrSerSerThrProSer 362
27059 GCTTGAACACACACGCGCGCGCTCGCGCTGCGCTGCGCTGCGCG 27010
363 LeuMetProAsnLeuGlnAsnPhePro..... 371
27009 TCGCGGAGACGCTTTCACCGCGCGCGCGGCAAGTCCGATCGC 26960
372 .....TyrSerGlnGlnProLeuSerThrGlyAlaP 382
26959 TGAAGACGCGGAGGCTTCGCGAGAGACGATGTACCCG.... 26914
382 heProAlaGlyLeuThrAspHisSerHisPheMetProLeuLeuAsnPro 398
|||||
26913 ..CCACCGAGGCTTGGTGCAGCTCGCCCGCTGCGACGCTGCGACGCG 26866
399 .....SerProThrAspAlaThrSerSerValAspThr...GlnAl 411
26865 AGTGCAGCGCCACGACGAGAGAGACGCGCTGATCCACACCGAGGT 26816
411 agLysAsnGlyLysProLeuGlnLysAspLysLeuProGlnAsnLeu 428
|||||
26815 CGGCGCGTGCAGGCGCCAGGAAATGATGACCGCTCGCGCGCAACCTGA 26766
428 eAspLeuSerLeuGlnSerLeuThrAlaLeuThrLeuGlnValGln 444
|||||
26765 GCATGTTGCCGCTGCGC..... 26749
445 lIleSerAsnThrValGlnGlnLeuLeuLeuSerLysAlaAlaValPro 461
|||||
26748 .....CTGTGACCCCGCGCGCTCATCGCGC 26723
461 nLysLysGlyValLysAsnLeuValSerArgThrProGlnGlnHisLys 478
|||||
26722 GAGTTCCTGACCGCGCTGCGATTCGTTGGGCGCGCGCCACACAC 26674
478 eArgHisGlySerProGlnGlySerGlyTyrSerAlaGlnProAla G 494
|||||

```

26673 .....ACGCCGCTGGGGCTCTCCGCGACGCCGACGGGTCTCTGG 26635  
494 yThrProLeuSerGluProSerSerThrProGlnSerThrHis...A 510  
26634 CCGCGCGCTCGATCGCTCCACCTCACTCCACACGACGCGCTGTG 26585  
510 laGluProGlnGluAlaAspTyrLeuSerGlySerGluAspProLeuGlu 526  
26584 CCGGTCCA.....GGCTCATCTCCCTCCGAGGG 26556  
527 ArgSerPheLeuTyrCysAsnGlnAla.....ArgGlySerProAla 541  
26555 AGA.....TGTGGAGAGACCGCGATCAAGCTCTCCACTCG 26518  
541 gValAsnSerAspSerLysAlaLysProGlnSerValSerThrCysSerV 558  
26517 CCGAGAAAGCCACCTCGGACCTAG.....GTCTGTCTCT 26481  
558 alThrSerProAspAspMetSerThrLysSerAspSerPheGlnSer 574  
26480 GTCCCTCTCCCGAGCGG...GGAGCGCGCTCTGCCCATTCACCGG 26434  
575 LeuHisGlySerLeuProLeu..... 581  
26433 TCGCGCGCACTCGGTGTGACACACCGCTCGGTCAACAGCTGCCA 26384  
582 ....AspSerPheSerLysPheValalaglyGluArgaspCysProArgL 597  
26383 GTAGACCTCCAGCTCTCGACCGCCCGCGGAGCGG..... 26347  
597 euLeuLeuSerAlaLeuAlaGlnGluAspLeuAlaSerGluLeuLeuGly 613  
26346 .....CAGGCGCACCCAGATGCGATCGGTCTCTGTCGCGCGC 26305  
614 .....LeuGlnGluAlaLeuGlyGluLysAlaAspLysAlaTrpAl 627  
26304 AACGAGCTCTCCGCGACGTCGCTGCTCTGACCTCCAGCGCTGGCT 26255  
627 aGluAlaProSerLeu.....V 633  
26254 CAGCAAGTATTCACACGCGCTCCAGGTCGGATGTCGACGCTAGCG 26205  
633 alLysAspSer.....SerLysProProPheSerLeuGluAsnHis 646  
26204 TCGCGACAGCGCATACCCAGCTCACCTGAAGCGCTTGGGATCTCC 26155  
647 SerAlaCysLeuAspSerValAlaLysSerAlaTrpProArg..... 660  
26154 ACGGCATCAGGAGTCCGAGCGCTCTCGCGAAGCTCTGTCGACGTC 26105  
661 .....ProGlyGluProGlnAlaLeuPro.... 668  
26104 GAGCGTGCCCTGCTGGTAAAGCCATCACCCGCGGACACGCGCGCA 26055  
669 .....AspSerLeu.....Gln 572  
26054 CAGCTCTAGACGCGGAGCGGTCTCCACACAGACGCGGTTCGCCAG 26005  
673 LeuAspLysGlyGlyAsnAlaLysAspPheSerProGlyLeuPheGluAs 689  
26004 CGTCCACAGCTGGCAGGCGGAG.....GAGA 25976  
689 pProSerValAlaPheAlaThr...ProAspProLysThrGly... 704  
25975 GGCCTCTTTCGGCAGGTACACGCCATCCAGAGCGGCTCGGCTCG 25926  
705 .....ProLeuSerPheGlyThrLysProThr..... 713  
25925 CGTCCGCGGCGGCTCCCGCATGGGCCCATTCATCTGGATGACACG 25876  
714 .....LeuGlyValProAla...ProAspProT 722  
25875 CGCTCGCTCGCGCGGTACGAGCAGCATTCATCCATCGCGCCAGCGCG 25826

722 hThrAlaAlaPheAspCysPheProAspThrThr.....AlaAla 735  
25825 ACTCGTGGCATTCGCCAGATCCCGATGCTCTGTTTCCCGCGCTCG 25776  
736 SerSerAlaAspSerAlaAsnProPheAlaTrpProGlnGluAsnLeuG1 752  
25775 CQTGCGAGCCATCCCTCCCTCCGCCACAGCCGCGATGCTCAGG 25726  
752 yAspAlaCysProArgTyrPheLeuHisProGlyGluLeuThrLysGlyL 769  
25725 GCGCGAGCCCTTCGCTTCGCT..... 25704  
769 euGlnGlnGlyLysAlaSerAspGlyLeuSerLysGlyAspThrHis 785  
25703 .....GGAGGT..... 25698  
786 GluAlaSerAlaCysLeuGlyPheGlnGlnGluAspProProGlyGluL 802  
25697 .....CCGCCAGCG..... 25689  
802 sValAlaSerLeuProGlyAspPheLysGlnGlnGluValGlyValL 819  
25689 ..... 25689  
819 ysGluGluAlaGlyLysLeuGlnCysProGlnValAlaLysAlaAsp 835  
25689 ..... 25689  
836 ArgTrpLeuGluAspSerArgHisCys.CysSerThrAlaAspPheGlyA 852  
25688 .....CGTCCAAAAGGCATTCCTCGCGGTAGCTGCCCTGGCGG 25648  
852 sLeuProLeuLeuProProThrSerArgLysGluAspLeuGluAlaGlu 868  
25647 AT.....CGAGCGAGAGAC..... 25634  
869 GluGluTyrSerSerLeuCysGluLeuLeuLysSerProGlnGluArgPr 885  
25633 .....GCCGAGCGCG 25623  
885 oGlyMetGlnAspProLeuSerProLysAlaProLeuIleCysThrLysG 902  
25622 AGGAAACAG.....TAGC 25609  
902 luGluValGluGluValLeuAspSerLysAlaGlyTrpGlySerProCys 918  
25608 AAGAGGTCCAGCGGTCTCCGGGTAAAG...GGTGTGACAGCACCATGC 25562  
919 .....HisLeuSerGlyGlu.SerValIleLeuLeuGlyProThrValG 933  
25561 CCGCTCCACCTTGGGCGCAACACC..... 25537  
933 lyThrGluSerLysValGlnSerTrpPheGluSerSerLeuSerHisMet 949  
25536 ..CGGCGGACCGACAGCGTCTGTGTGGCGACAGCCGCTGCTGAGC 25489  
950 LysPro.....glyGluGluGlyProAs 957  
25488 AGACCCGCGGCTGCACTACCCCGCGCGGCGGTCTGACGCGCGCCAA 25439  
957 pGlyGluArgAlaProGlyAspSerThrThrSerAspAlaSer..... 971  
25438 GAGCGCGCGCATCGCTCGGCATCGGCCAGCTCGACGCGCGCGCGTGA 25389  
972 .....LeuAlaGlnLys.ProAsnLysProAlaValProGluAl 984  
25388 CCGCGCGCGCTCGGCTTCAGCGCTCGATGCGCGCATGCGCGCGCGC 25339  
984 pProIleAlaLysLys...GluProValProArgGlyLysSerLeuArgS 1000  
25338 ACCTCTGCGGTGATCTCGGCCCATTCCTCGCGTCCGCGCAATCCGTG 25289

[illegible]

```

24459 GGGTCGAGACCTTCGG.....CCGAGCGCCT... 24474
|||||
1255 LysLYSArgAsnLeuValLeuArgHisGlySerSerSerSerSerAsnI 1251
|||||
24473 .....CGATCCAACTATACCCCTCGCGCCACAGAGATAGTGGGAGCGCC 24430
|||||
1251 AsnAlaMetGlyLysMetGlyArgArgArgGlyLeuArgValProPro 1267
|||||
24429 ACCCGCGCGCCACGCCGTGGAGAGACG.....CCGGCCA 24395
|||||
1268 .....ProSerSerArg.....GlyCysLeuLeuProArg...Lys 1278
|||||
24394 GCTGACCAAGCGCGCGCGGCCGCCAACAGCTGCCACAGCGCTCGAGACCC 24345
|||||
1279 ProSerProProArgAla..MetAlaSerLeuProGlnSerHisSer 1294
|||||
24344 CCGGAGCTCTCCGCGCGCGCGCCGACACAGCAGCGAGCGTGGCGTCC 24295
|||||
1294 ArgArgArgProProPheLeuProGlnAlaArgLeuSerAlaAlaPheGln 1311
|||||
24294 GCGTCCGCGAGCGAGCTGGCAACAGCCCGACAGCAGAGTGGGTGGCGGCC 24245
|||||
1311 LysAlaMetLysThrLysValLeuProProArgGlyGlyArgGlyLys 1327
|||||
24244 CACTTCGACGAACGCTCC.....CCGCGCGCGCTTGT..... 24213
|||||
1328 LeuGlnAlaIleValGlnLysIleThrSerProSerLeuLysLysPheAl 1344
|||||
24212 .....GCAGCGCTTCACCCCGCGCGAGAA 24187
|||||
1344 AcGlyLysAlaProGlyLysAlaSerProGlyLysProLeuSerProSerLeuS 1361
|||||
24186 C.....GCACCGCTCCCGCCAGCTGCCAGCCACCAAGTACCCGCGCGCT 24143
|||||
1361 ArgAspLysAspArgGlyLeuLysGlyAlaGlyLysSerProValGlyVal 1377
|||||
24142 CAGCTCGTCCGAGACACCTCCCGCTCAGGTTCTCTACACAGC..... 24099
|||||
1378 GlnGlnLysLeuValAsnValGlyThrGlyGlnLysLeuProThrHisG 1394
|||||
24098 .....AAAGCGTTGGCGCGCGGT.....ACGTCACC 24073
|||||
1394 ValAspProLeuGlyAspArgAsnProThrAsnArgSerLeuLysGlyLysI 1411
|||||
24072 GACGCGCCACCGCGCCCAACTCTCCACAGCTCGATCCATCGAGCGCGA 24022
|||||
1411 eumEtAsnSerLysLysLeuSerSerThrAspCysPheLysThrGlnAla 1427
|||||
24022 GTGGAGAGCGTGGAGAAATGACAGCCGCTGTGTGCGGCGCGCGCGCG 23973
|||||
1428 PheThrSerProGlnAlaLeuGlnProGlyLysThrAlaLeuAlaProLy 1444
|||||
23972 GCAACCGCGCGCGAGTACCGTCGACACG.....CTTGCTCCACG 23933
|||||
1444 sLYsArgSerArgLysGlyArgAlaGlyAlaHisGlyLeuSerLysGlyP 1461
|||||
23934 CCGCGAGTACACACCT.....GCTCGGCGCCATTCAGCCGCGCGATGA 23891
|||||
1461 rOLeuGlnLysArg.....ProTYrLeuGlyProAla 1471
|||||
23890 CACCGACCGCGCTGGGCGCCACCGCGCGCCACCTCGCGCTCCGCGC 23844
|||||
1472 LeuLeuLeuThrProArgAspArgAlaSerGlyThrGlnLysAlaSerG 1488
|||||
23840 CTCGAGAGACACACAGCGCGCGCGCGAGAGCGCCCTGCATACAGCC... 23792
|||||
1488 uAspAsnSerGlyGlyGlyLysLysProLysMetGlnGlnLeuGlyP 1505
|||||
23792 .....GCCCGCGCGCGCCAGACCGACCCACCCACCTTC 23755
|||||
1505 rOLaSerGlnProProGlnGlyArgProCysGlnProGlnThrArgAla 1521
|||||

```

```
23758 CAGCGAGAACACCCCGCCACGACGCG...CCACGACTCCCGGATGC 23712
1522 GlnLysGlnProGly.....HisThrAsnTyrSerSery 1533
23711 TATGCCAACAGAGAGTCGGCTACGCCGCCACGACGCCACAGCGCC 23662
1533 rSerLysArgLysLeuThrArgGlyArgAlaLysAsnThrThrSers 1550
23661 GTCAGCGGTACTCCACCGCGAAGACGCGGTGGTGAACGCGTCT. 23613
1550 erProCysLysGlyArgAlaLysArgArgGlnGlnValLeuPro 1566
23612 .....GGTCGAGCAACAGACTCGCGCTCC.....CCG 23583
1567 LeuAspProAlaGluProGluLeuLeuArgLeuLysTyrLysSerCysLy 1583
23582 CTTCCGCCACATCACT.....CGCGCAGCGGGGTCACG 23545
1583 sArgLeuArgSerSerArgThrProAlaPheSerProPheValArgV 1600
23544 TCCTGGTCAACAGCGCCACGACCGCTCGAACCGCTCC..... 23505
1600 alGluLysArgAspAlaPheThrThrLeuCysThrValValAsnSerPro 1616
23504 .....GGACGCTGGCCACGCGCGCAAGCCGCC 23475
1617 GlyAspAlaPro.....LysProHisArg.....LysProSe 1627
23474 GGCCATGTCGGCGCTGTCGGCGCTGTCGGTGAACAGAACGCCACG 23425
1627 rSerSerAlaSerSerSerSerSerSerPheSerLeuAspAlaA 1644
23424 TTGCGCGGAGGAGCTCGCGATGCAGC..... 23397
1644 laGlyAlaSerLeuAlaThrLeuProGlyLysSerLeuLysProArg 1660
23396 .....CGCGCGCCCGCGCGCTGTCCTGCGCGCGG 23361
1661 ProSerLeuProLeuSerThrMetHisLeuGlyProValValSerLy 1677
23360 CGGAGAGCGCCGACGAGCCCTCGCGGAGCTCAGCGCCACCGGAGC 23311
1677 sAlaLeuSerThrSerCysLeuValCysCysLeuCysGlnAsnProAla. 1693
23310 CGGTGGCTCATCGCGCTC.....CGCTGTCGCCAGCT 23276
1694 .....AsnPheLysAspLeuGlyAspLeuCysGly..... 1703
23275 GAACGCCAGCTCCCGACCGGAGCTCCGGTGCATGTCCAGTGTCTCC 23226
1704 .....ProTyrTyrProGluHisCysLeuProLysLysLysProLy 1717
23225 GCAGCGCGCGC.....CTGCGCATCGAGCGCCCTGCTCTTGCOC 23182
1717 sLeuLysGlyLysValArgProGluGlyThrCysGluGluAlaSerLeuP 1734
23181 GACAGCACCAAGAGTCCCGAGCGCTCCGCGCGGAGGCCACAGCTC 23132
1734 roLeuGluArgThrLeuLysGlyProGlu..... 1743
23131 CACCGCGCGGCTCTTCCAGCACCATGCGGTTCCTTCGCTCATCC 23082
1744 .....CysAlaAlaAlaAla 1748
23081 CGAAGAGCTCACCCCGGAGCGCGGGGTGCTCGTGGCGGCGCAGCG 23032
1748 aThrAlaGlyLysProProArg 1755
23031 ACCGGCTCGTCCCAACGCA 23010
```

seq\_name: /cgn2\_6/ptdata/2/ina/6B\_COMB.seq:US-09-443-501A-2

seq\_documentation\_block:

```
; Sequence 2, Application US/09443501A
; Patent No. 630342
; GENERAL INFORMATION:
; APPLICANT: Kosan Biosciences, Inc.
; APPLICANT: Julien, Bryan
; APPLICANT: Katz, Leonard
; APPLICANT: Khosla, Chaitan
; APPLICANT: Tang, Li
; APPLICANT: Ziermann, Rainer
; TITLE OF INVENTION: Recombinant Methods and Materials for Producing
; FILE OF INVENTION: Epothilone and Epothilone Derivatives
; FILE REFERENCE: 30062-20031.00
; CURRENT APPLICATION NUMBER: US/09/443.501A
; CURRENT FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: US 60/130,560
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: US 60/122,620
; PRIOR FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: US 60/119,386
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: US 60/109,401
; PRIOR FILING DATE: 1998-11-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 71989
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct
; US-09-443-501A-2
```

alignment\_scores:
Quality: 462.50 Length: 2116
Ratio: 0.518 Gaps: 119
Percent Similarity: 42.202 Percent Identity: 22.590

alignment\_block:

US-09-508-821B-6 x US-09-443-501A-2/rev ..

Align seg 1/1 to reverse of: US-09-443-501A-2 from: 1 to: 71989

```
33. ProSerGlnAlaGlyLeuSerCysAspArgGlnArgLeuAlaLysAs 49
|||||:|||||: ||| :|||||: |||
22992 CCTTCCGCTCAGTTCCTCA...CCAGCGAAGACGCTTGGCGCGGTAC 22946
49. pTyrTyrAsnProGlnProTyrProSerTyrGluGlyAlaGlyThrP 66
|||||: ||| |||: |||: |||: |||
22945 G.....TCACGCGAGCGCGCCACCGCGGACTCTCCAGCATCGTTC 22902
66 roSerGlyThr.....AlaAlaAlaAlaAla 75
|||||: |||: |||: |||: |||
22901 CATCAGCGCGAGTGGATGCGTGGAGACATCGCGCTTGGTGGCA 22852
76 AspLysTyrHisArgGlySerLysAlaLeuProThrGlnGlyLeuL 92
|||||: ||| |||: |||: |||: |||
22851 CGCGCGCGCGCGAACCAGCGCGGATCGCTGCGACCGCTTGTCCACG 22802
92 nglyArg...ProAlaPheProGlyTyrGlyValGlnAspSerSerPro 108
|||||: ||| |||: |||: |||: |||
22801 CCGCGGATCACCCTCTCGCGGCCATTGACCGCGCGATCGACACCA 22752
108 yProGlyArgTyrAlaGlyGluGluSerLeuGlnAlaTrpGlyAlaPro 124
|||||: ||| |||: |||: |||: |||
22751 CGCGG.....CGTGGCGCGCCACCGCGC 22729
125 GlnProProPro.....GlnProGln 132
|||||: |||: |||: |||: |||
22728 CGGCCACTTCCGCTCGCGGCTCGAGCGGACCATCGCGCGCGCGCC 22679
132 nProLeuProAlaGlyValAla.....LysTyrAspGluAsnL 145
|||||: |||: |||: |||: |||
```

22678 GAGAGCCCTGATCAGCGCCGCGCGCCAGAGCTCAGCCCATC 22629  
 145 eumetylserThrAlaValProPro..SerArgGlnTyrAlaGlnGln 161  
 22628 TTCACGCGAGAACACC...CGCCACGCGAGCGCCAGCAGCTCCCGC 22582  
 161 yAlaGlnValProPheArgThrHisSerLeuHisValGlnGlnProPhe 178  
 22581 CGCT.....ATGCCAGCCACGAGCTCGAGCTTCAGCCGCCAGAC 22541  
 178 roProGlnGln..ProLeuAlaTyrProLysLeuGlnArgGlnLysLeu 194  
 22540 CGCCACAGCGCCGTCAGCGGTCTCCAGCTGAGAGAGCGCGGTGGGT 22491  
 194 nasAspIleAlaSer.....ProLeuProPheProGlnLysThr 208  
 22490 GAACGCGCTGTGTGTCAGCAACAGACTGCGCGCCGCGCTCCGCC 22441  
 208 AspPheProGlnHis.....SerGlnSerPheProThrSerSerThr 221  
 22440 ACATCAGCTCCGACAGCGCGCGGTCCAGCTCCGCGTCAAGACGCCACG 22391  
 222 Tyr..SerSerSerValGlnGlnLysGlnGlnLysSerTyrLys 237  
 22390 CACGCGTCAACGCGCTCTCGAACGCTGGCCAGCGCGCCAGAGCCCGC 22341  
 237 sSerCysThrAlaProThrAlaGlnProHisAspArgProLeuThrAla 254  
 22340 GCC.....CATGCCGCGCTGCGCGCC 22316  
 254 erSerSerLeuAlaProGlnArgValGlnAsnLeuHisAlaTyrGln 270  
 22315 TGTC.....GGTGAACAGAGAACCGAGCTGCCGCG... 22284  
 271 SerGlyArgLeuSerTyrAspGlnGlnGlnGlnGlnGlnGlnGln 285  
 22283 .....CGACGAGCTCGGATGACGCGCGCGCGCGCGCGCTGTC 22240  
 286 .....GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 298  
 22239 CCGCCGCGAGCGCGAGAACGCGCGCGCGCGCGCGCGCGCGCGCGC 22190  
 298 lAlaGlnGlnThrLeuHisTyrGlnAsnLeuAlaTyrGln..... 312  
 22189 GCCACGCGAGCGCGGTGTCATCGCGCTGCGCGCGCGCGCGCGCGCA 22140  
 313 ..HisTyrGlnGln..GlnGlnGlnGlnTyrCysGlnProAspAla... 326  
 22139 CGCCAGCTCCGCGAGCGCGAGCTCGGAGTGCATGT...CCAGTCTCGC 22093  
 327 .....AlaValArgThrProGlnGlnTyrTyrGlnThrPh 338  
 22092 GCAAGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 22043  
 338 eSerProSerSerSerHisSerProAlaArgSerValGlyArgSerPro 355  
 22042 ACCAAAGCTCCGCGCGCTCCGCGCGCGCGCGCGCGCGCGCGCGC 21994  
 355 erTyrSerSerThrProSerProLeu...MetProAsnLeuLysAspHe 370  
 21993 .....CGGCGCTCTTCCAGCACACACATGCGGCTGCTC 21958  
 371 ProTyrSerGln.....GlnProLeuSerThrGlyAlaPheProAlaG 385  
 21957 CGCATCCCGAGAGAGCTCAGCCCGCGAGCGCGAGCGCGGTGCGC 21908  
 385 yIleThrAspHisSerHisPheMetProLeuLysAsnProSerProThr 402  
 21907 GCCACG.....GCACGCGCTCGCGCGCGCGCGCGCGCGCGCGC 21873  
 402 sPalaThrSerSerValAspThrGlnAlaGlnAsnCysLysProLeuGln 418  
 21872 CAGCGCTCCCGCAG.....GCAAGAGCGGTGTGAGCTGCGCCA..... 21835

419 LysAspLysLeuProGlnAsnLeuSerAspLeu..... 430  
 21834 .....GTCGCGCTGGGCTGGATTTTCCTTCTTGGCCAGCGCAGC 21794  
 431 .....SerLeu..GlnSerLeuThrAlaLeuThrGlnVal 442  
 21793 ACCGCTTGACAGACCGCGCGAGCCCGCGCGCGCTCCATGTGCCAAG 21744  
 443 GlnAsnIleSerAsnThrValGlnGlnLeuLeuSerLysAlaAlaVal 459  
 21743 GTTGCGCTTGCGCGCTCCAGATCAGCGGTGCGCGAGCGCGGT 21694  
 459 lPro.....GlnLysLysGlyValLysAsnLeuValSerArgT 472  
 21693 GCCGTACAGCTGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGC 21644  
 472 hrPro.....GlnGlnHis... 476  
 21643 GTCCGCTCCGCTGCGATTCAGAAATCGAGCTCGCGCGAGACACGCC 21594  
 477 .....LysSerG 479  
 21593 TGCGTGCCAGCGCTGCGGTACAGCGCTCTCGGCGAGGCGCGTGG 21544  
 479 nHisCysSerPro..... 483  
 21543 GCACTGACCGCGCTGCGCGCTGCGCGCTGATGATGCGCGCGCGC 21494  
 484 .....GlnGlySerGlyTyrSer.....AlaGlnPro..Al 493  
 21493 ATCAGCGCGAGATGGGTGCGGTGCGCGCGCGCGCGCGAGAGCGGT 21444  
 493 aGlyThrProLeuSerGlnPro.....ProSerSerThrP 505  
 21443 GAGACCGACAGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGC 21394  
 505 roGlnSerThr...HisAlaGlnProGlnGlnAlaAspTyrLeuSerGly 520  
 21393 CCGAGAGCTCTTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 21344  
 521 SerGlnAspProLeu.....GlnArgSerPheLeuTyrCysAsnGlnAl 535  
 21343 AGCAGCGCGAAGCTTCTGCGCGAGACAGCATGTG..... 21308  
 535 aArgGlySerProAlaArgValAsnSerAspSerLysAlaLysProGln 552  
 21307 .....ACCAGCGCGAGCGCTGCTGCGCGCTGCGCGCGCGCGC 21262  
 552 erValSerThrCysSerValThrSerProAspAsp.....MetSerThr 566  
 21261 GCGAGCGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 21212  
 567 .....LysSerAspSerPheGlnSe 574  
 21211 GCCAGGTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGC 21162  
 574 rLeuHisGlySerLeuProLeuAspSerPheSerLysPheValAlaGly 591  
 21161 AACGTGAGCATGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 21112  
 591 lArgAspCysProArgLeuLeuLeuSer..... 600  
 21111 CGAGGTCTGCGCGCGCTGCGCGATATGTGTTGGGCGCGCGCGCGC 21062  
 601 .....AlaLeuAlaGlnLys 606  
 21061 ACAGCGGTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGC 21012  
 606 pLeuAlaSerGlnLeuLeuGly..... 614  
 21011 GATGCGCTCCGAGTCACTTCCAGCAGCAGAGCGCGCTGTGGGGTCCAGC 20962

614 euGlnGluAlaIleGlyGluLysAlaAspLysAlaTrpAlaGluAlaPro 630  
 20961 TCATCGCTCCCGAGCGAG.....ATGCGAAGACCGCGCA 20924  
 631 SerLeuValLysAspSer...SerLysProPheSerLeuGluAsnHi 646  
 20923 TGGACGTCCTCCACTCGCGAGAACCAACCCTGGCAGG..... 20982  
 646 sSerAlaCysLeuAspSerValAlaLysSerAlaTrpProArgProGlyG 663  
 20881 .TAGGTCTGTCTC.....GGAGCTCTCCGAGCGCGGGC 20848  
 663 luProGluAlaLeuProAspSerLeuGlnLeuAspLysGlyAsnAla 679  
 20847 CGGCGCCCTCTGCCCCA.....TTCACCGGTGCGCGCGCACTCGGTG 20804  
 680 LysAspPheSerProGlyLeuPheGluAspProSerValAlaPheAlaTh 696  
 20803 .....CTGACCACCAACGCTCGGTCAACAGCTGCCA 20772  
 696 rProAspProLysThrThrGlyProLeuSerPheCylThrLysProTr 713  
 20771 GTAGACATCCAGTCTCGACCCGCGCGGAGCGGACGCGCGCGCACCA 20722  
 713 hrLeuGlyValProAlaPro.AspProThrThrAlaAlaPhe.....As 727  
 20721 CGATGGCATCGGCTCTCTGTCTCGCGGCAACCGAACCTTCGACGTG 20672  
 727 pCysPheProAspThrThrAlaAlaSerSerAlaAspSerAlaAspProp 744  
 20671 GTGCGTCTCTGCAGCTCCAGCGCTGCTCGACGA.....AGTATTCACCA 20625  
 744 heaLarProGluGluAsnLeuGlyAspAlaCysPro.....ArgTrp 758  
 20624 CGCTCCAGCTCGGATGTCCAGCGCCAGCTCGCGACAGCGGCAATAC 20575  
 759 GlyLeuHisProGlyGluLeuThrLysGlyLeuGluGlnGlyLysAl 775  
 20574 CCAGCTCACCTCGAAGCCGTTTGGGATCTCCACAGCATCAGGG..... 20530  
 775 aserAspGlyLeuSerLysGlyAspThrHisGluAlaSerAlaCysLeuG 792  
 20529 .AGTCGAGCCCTGCTCGCGAGACCTCTGTCTGCAGTCGACGTGCTGTT 20481  
 792 lyPheGlnGluAspProProGlyGluLysValAlaSerLeuProGlu 808  
 20480 GT.....CGGTAAAGCCCATCACCCGCG 20458  
 808 yAspPheLysGlnGluValGlyGlyValLysGlnGluAlaCly..... 823  
 20457 CGAACCCGCGGCACAAGCTC.....GTAGAGCGCGAGCGGTCTCC 20414  
 823 ..... 823  
 20413 ACAACAGACGCGTTGCGCCAGCGCTCTACAGCTGGCACGCGCGAGNGCA 20364  
 824 .....GlyLeuLeuLysCysProGluValAlaLysAlaAs 835  
 20363 GGCGCGTTCGTGACAGTTACACCGCATCCACAGGCGCGCTCGCTGCT 20316  
 835 pArgTrpLeuGluAspSerArgHisCysCysSerThrAlaAspPheCylA 852  
 20315 .....CGGTCTCGCGGACCGCTCCGCGATGGCGC 20285  
 852 spLeuProLeu.....LeuProPro 858  
 20284 CAATCCATCTGATGATCCACGCGTGTGCTCGCGCGTACCGAGCAGCA 20235  
 859 ThrSerArgLysGluAspLeuGluAlaGluGluLysThrSerSerLeuCy 875  
 20234 TTCATCTGCGCCCGAGCGACGACTCGTCTGCGATCGCGCATGATCCCGAG 20185  
 875 sGluLeuLeuGlySerProGluGlnArgProGlyMetGlnAspProLeu 892

20184 C.....CTCATGTTCCGGCGCTGCCTTCAGAGCCATCCCCCCTCC 20141  
992 ePrPolyAlaProT...|||||...:::||||||  
|||||:|||||  
20140 GCCACGCCGCCAGCGATGTCTAGCGCGGAGCCCTGCATTGCATGCAGG 20091  
897 .....LeuLieCyS.....899  
20090 GAGGTCCGCCAGCGCTTCAAAAGGCATTGCCTGCGCGTAGCTGCCCT 20041  
900 .....ThrLysGlusIuV 904  
|||||:|||||  
20040 GGCGATCGAGCGGAAGACGCCGACGCGAGAAAACAGTACGAAGAG 19991  
904 aIGlUGluValLeuAspSerLysalaGIyTrpGLySerProCyS.....918  
|||||:|||||:||||| ||| |||||  
19990 TCAGCGCTCTCGCGGTAG..GGTGCGACACCACATGCCCTTC 19944  
919 HisLeuSerGIyGlu\_SerValIIeLeuLeuglyProThrValGIYThrG 935  
|||||:|||||: |||  
19843 CACCTGGGGCGCAACACC.....CGG 19821  
935 luSerLysValGlnSerTrpPheGlnSerSerLeuSerHisMetLysPro 951  
|||||:|||||: ||||| ||| |||||  
19820 CGAGCGACCGGCTCTCTGGTGCGCACGAGCCCTCTGCAGCAGACC 19871  
952 .....GlyGluGluGlyProAspGIyG 959  
|||||: |||: |||:  
19870 GCGGCTGCAGACCCCAGCGGGCTCGAGCGCGCCAGAGCCG 19821  
959 uARgAlaProGIyASerThrThrSerAspAlaser.....971  
|||||: ||||: ||||| |||||  
19820 CGCATGCTTTCGCATCGGCCACGCTGAGCGCGCCGCGTACCCGG 19771  
972 .....LeuAlaGlnLys\_ProAsnLysProAlaValProGluAlaProLI 986  
|||||: ||||: ||||| ||||| |||||  
19770 CGCTCGGCTCTCAGGCTCGATGCCGCATNGCGCGCGGCACTCT 19721  
986 eaLalysLys...GluProValProArgLyLysSerLeuArgSerArgA 1002  
||| |||||: |||||: |||||: |||||: |||||: |||||:  
19720 GCGGCTGATCTCGGCCCATTCCTCGGTCGGCAATCGTGGCGCT 19671  
1002 rgValHisArgLyLeuPro.....1008  
:|||| |||  
19670 GATGACAAGAAGTGCOCGCGCCGCTCCACCAACACGCGGCAACGA 19621  
1009 .....GluAlaGlnAspSerProCySArgAlaProValLe 1020  
19620 GAGGCCAAGGCGCCACGCCACCCGTACCAAGTAAGTCCCT.... 19576  
1020 uProLysAspLeuLeuLeuProGluSerCysThrGly.....P 1033  
19575 .....CCGACAGACGACCGGCGCTGCGTTCC 19545  
1033 roProGlnGlnMetGluGly.....AlaGlyAlaProGIyArg 1046  
||||| |||||: |||||: |||||: |||||: |||||: |||||:  
19544 CTCGGTGGGCGGCACAGCGCGGTGCGCGCGCGCCCTTCGCG... 19498  
1047 GlyAlaSerGIyGluLeuProArgMetCysThrArgSerLeuThrAlaLe 1063  
19497 .....GGATGCCACTGATCCTCGGCTCGCGGAAGCA.. 19462  
1063 userGluProArgThrProGlyPro.....1071  
||| |||||  
19461 .....GCTCGGCACACGAGGCTCGAACCTCGTGGGCTCTCTCCG 19419  
1072 ..ProGlyLeuThrThrPro.....1078  
||||| :|||:  
19418 ATCCAGTCCAGAGGCCGCCCCAGAGCGGGAATGCTCAGAGCGCGCA 19369  
1079 .....AlaProAspLys.....1083  
|||||: |||||:



19368 CCGGGCCATACCCGACGCGCCCTGACAGGGGCAAGCGTACGCTCG 19319  
1084 .....LeuGlyLysGlnArgAlaIlePheLys 1094  
19318 CCGCCACCGGTGACAGGCCCTCGGCTACAGATCCAGAGCGGGGTAGCG 19269  
1094 ergLysArgValGlyLysProSerProLysAlaIleSerProSer 1110  
19268 CCGGCCCG.....TGCACACGCTGATCAGCGAGCAGCAGCGCG 19228  
1111 .....As 1111  
19227 CCGTGGCAAGATGGGTACTTGCACACTCTTCGCGCCATGCCCCCGC 19178  
1111 nProAlaIleuProValAlaSerAspSerProMetGlySerLysT 1128  
19177 TCACAGACGCGCTCCAGACCCAGAGTACAGACCCCGGACGCTT 19128  
1128 hLysGlnThrAspSerProSerThrProGlyLysAspGlnArgSerMet 1144  
19127 GCGGCGACCGA...GGGCTGGGTACCTGCTGCGGACGCGG..... 19087  
1145 IleuArgSerArgThrLysThrGlnIlePheHisSerLysArg 1161  
19086 .....AGGCTCG 19079  
1161 gArgProSerGlnGlyArgIleProAsnCysArgAlaThrLys..... 1175  
19078 GCGGCGCATGAGCAGCGGCAACATCCCTGCGAGGAGGCGCGC 19029  
1176 .....LysLeuAspAsnSerHisLeuProAlaThrPhe 1187  
19028 CCGGCGCGCTCCCGACTCCACCCGCTGGCGCAGCAGCGACCGCCG 18979  
1188 Lys.....ValSerSerProGlnLysGlnArgValSerLys 1202  
18978 CCGACCGGCTCGCGCGATCA.....CGAGACGAGG 18941  
1202 gAlaArgValProLysProGlyAlaGlySerLysSerAspArgPro 1218  
18940 ATCTCGGCGAGTCCACCGGTAGAACCACTGCGCAGCGATGGGCGC 18891  
1219 .....LeuHisAlaLeuLysArgLysSerLysPheMetAlaProValPro 1233  
18890 CGTGGCTCGAGCCCTCGG.....CGGCGCT 18862  
1234 ThrLysArgAsnLeuValLeuArgHisGlySerSerSerSer 1250  
18861 .....CGATCCAGTACCGCTCGCGCGCAGCGATAGTGGGAGC 18821  
1250 nAlaSerLysMetGlyMetGlyArgArgArgGlyLeuArgValPro 1267  
18820 GGCACCGCGCGCCAGCGCTGGGAGAGAGC..... 18790  
1267 roProSerSer.....ArgGlyCysLeuLeuPro 1276  
18789 ..CCGCGACGTACCGACCGCGCAGCGCCACACCTCGCCAGCGCT 18742  
1277 Arg...LysProSerProProArgAla...MetAlaSerLeuProGln 1291  
18741 CAGACACCCCGCAGCTCTCTCGCGCGCGCGCAGCAGCAGCAGCAGC 18692  
1291 rSerHisProArgArgProProPheLeuProGlnAlaArgLeuSerLys 1308  
18691 GCGGCTCGGCTCGCGCAGCGAGCTGGCAGACAGCGCAGCAGCAGC 18642  
1308 laphGlnGlyAlaMetLysThrLysValLeuProProArgLysGlyArg 1324  
18641 CT...TCGGGCCACTTCAGAGAGAGTGC.....CCGCGCGCTCGCT 18601  
1325 GlyLeuLysLeuGlnAlaIleValGlnLysIleThrSerProSerLeu 1341  
18600 .....CGAGCGCTTCACCCCG 18584

1341 sLysPheLysLysAlaProGlyAlaSerProGlyLysProLeuSer 1358  
18583 TCGGGAAGC.....GCAGCGCTCCCGGACGTGCGGCCAGTGGC 18540  
1358 roSerLeuSerAspLysAspArgGlyLeuGlyAlaGlyLysPro 1374  
18539 CCGCGCGCTCAGCTCGCTCGGACACCTCCCGCTCAGAGTGGTACCA 18450  
1375 ValGlyValGlnGlnGlyLeuValAsnValGlyThrGlyGlnLysLeu 1391  
18489 GCG.....AAAGCTGGCGCGCT..... 18469  
1391 oThrSerGlyAlaAspProLeuCysArgAsnProThrAsnArgSerLeu 1408  
18468 .AGCTACCGACGCGCGCACCGCGGAACTCTCCAGATCGGTCCAT 18420  
1408 yseGlyLysLeuMetAsnSerLysLysSerSerThrAspCysPheLys 1424  
18419 CACCGCGAGGAGAACCGCTGGGAGACATGACCGCTGTGGCGACGC 18370  
1425 ThrGlnAlaPheThrSerProGlnAlaLeuGlnProGlyGlyThrAla 1441  
18369 CCGCGCGCGCAGAACCGCGCGGATCGCTGACCG.....CT 18332  
1441 uAlaProLysLysArgSerArgLysGlyArgAlaGlyAlaHis..... 1455  
18331 TGCTCCAGCGCGGATCCACCT.....GCTCGGCGCTTGACCGC 18288  
1456 .....GlyLeuSerLysGlyProLeuGlnLysArgProTyLeu 1468  
18287 CCGGATCAGACACCGCGCGTGGGCGGCGCAGCGCGCACCTCGC 18238  
1468 GlyProAlaLeuLeuThrProArgAspArgAlaSerGlyThrGln 1485  
18237 CTCCCGCGCTCCGAGGACACACCGCGCGCGCGCGCGCGCTCGC 18188  
1485 yAlaSerGlnAspAsnSerGlyGlyGlyLysLysProLysMetGln 1502  
18187 ATCAGC.....GCCCGCGCGCGCAGCAGAGCTTAC 18156  
1502 LuLeuGlyProAlaSerGlnProProGlnGlyArgProCysGlnPro 1518  
18155 CCGATCTCCAGCGAGACACCGCGCGCAGCGCGC...CCACCACT 18109  
1519 ThrArgAlaGlnLysGlnProGly.....HisThrAsnTy 1530  
18108 CCGGATGCTATGCCAACCAGAGCTCGGCTTACGCCCGACGACCGC 18059  
1530 rSerSerLysLysArgLysArgLeuThrArgGlyArgAlaLysAsn 1547  
18058 CAGAGCGCGTACGCGTACCGACCGCGAGAGCGCGGTGGGTGA 18009  
1547 hThrSerSerProCysLysGlyArgAlaLysArgArgGlnGln 1563  
18008 CCGCGCTGCTCCAGC.....ACAGACACTCGCGCTCC..... 17974  
1564 ValLeuProLeuAspProAlaGlnProGlnIleArgLeuLysTyIle 1580  
17973 .....CGGCGCGCGCACATCCT.....CGGCGCGCG 17942  
1580 rSerCysLysArgLeuArgSerAspSerArgThrProAlaPheSerPro 1597  
17941 CCGTCCAGCTCCCGGTGAAACAGCAGCAGCAGCGTGAACCTTCC 17893  
1597 heValArgValGlnLysArgAspAlaPheThrThrIleCysThrVal 1613  
17892 .....GGAAGCTGGCGCGCGCGC 17872  
1614 AsnSerProGlyAspAlaPro.....LysProHisArg..... 1624  
17871 AGAGCCCGCGCGCATGCGCGGTGCGCGCGCTCGGTGAGACGC 17822

```

1625 .LysProSerSerSerAlaSerSerSerSerSerSerSerSerPheSerL 1641
17821 AACGCCAGCTTGGCGGAGAGAGCTCGCATGACG..... 17785
1641 euAspAlaAlaGlyAlaSerLeuAlaThrLeuProGlySerIleLeu 1657
17784 .....GCGCGCGCCCGCGCGCTCTGCCCT 17758
1658 GlnProArgProSerLeuProLeuSerSerThrMetHisLeuGlyProVa 1674
17757 GCGCACGCGCGGAGAGCGCGGAGAGCGCGCTCGCGGAGCTCAGCGGC 17708
1674 lValSerIyAlaLeuSerThrSerCysLeuValCysCysLeuAla 1691
17707 ACCGGAGCGCGTGGGTATCGCGCTC.....CGGTCT 17673
1691 snProAla.....AsnPhelyAspLeuGlyAspLeuCys 1702
17672 GCCAGGCTGAGCGCACGCTCCCGGAGCGGAGCTCGGGTGCATGCCA 17623
1703 Gly.....ProTyrTyrProGluHisCysLeuProLysLy 1714
17622 GGTGCTCGCGAGCGCGCG.....CCTGCGCTCGAGCGCGCTCG 17579
1714 slsProLysLeuLysLysValArgProGluGlyThrCysGluGluA 1731
17578 CTCTGCGCGGACGACCAAAAGCTCCGCGGAGCGCTCGCGCGGCGG 17529
1731 laSerLeuProGluArgThrLeuLysGlyProLys..... 1743
17528 CCACAGCTCCACCGCGCGGCGCTCTCCAGCACACCATGCGGTTC 17479
1744 .....CysAl 1745
17478 CGTCATCCCGAGAGAGCTCACCCCGGAGCGGAGCGGTCTCGCGC 17429
1745 alaAlaAlaThrAlaGlyProProArg 1755
17428 GGCACGCGCGCGCTCGGTTCGCCAAGCGA 17398

```

seq\_name: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-09-105-537-1

# seq\_documentation\_block:

```

; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.4380S1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-1

```

## alignment\_scores:

```

Quality: 454.50      Length: 2052
Ratio: 0.521         Gaps: 117
Percent Similarity: 42.544      Percent Identity: 23.635

```

## alignment\_block:

```

US-09-508-821B-6 x US-09-105-537-1
Align seg 1/1 to: US-09-105-537-1 from: 1 to: 15872

```

```

31 ArgGlnProSerGlnAlaGlyLeuSerCysAspArgGlnArgLeuAl 47
639 AGAGCCCGCTGACGAGGAGCGCTTCGGTGAGCTCCCGCG..... 680
47 aLysAspTyrTyrAsnProGlnProTyrProSerTyrGlyGlyAlaG 64
681 .....ACGGCAGCGCTACACCTTCGAGCGCGGG..... 710
64 lyThrProSerGlyThrAlaAlaValAlaAlaAspLysTyrHisArg 80
711 .....CCACGAGATTCTCGGGGCGAGGCGCGGAGTGC..... 746
81 GlySerLysAlaLeuProThrGlnGlnGlyLeuGlnLysArgProAlaPh 97
746 ..... 746
97 eProGlyTyrGlyValGlnAspSerSerProTyrProGlyArgTyrAlaG 114
747 ....TCGTACTCAAGCGGCTCTCCGCGCGCTCGCGAGCGGAGCGGTGT 792
114 lyGluGluSerLeuGlnAlaThrGlyAlaProGlnPro..... 126
793 CCACGGCGTCA.....TCGGCGCGAGCGCGTCAACACGAGCG 830
127 ProProGlnProGlnProLeuProAlaGlyValAlaLysTyrAspG1 143
831 GAGCACCCCGGGTCTCACGTCGCCGAGCAGGG..... 863
143 uAsnLeuMetLysLysThrAlaValProSerArgGlnTyrAlaGluG 160
864 .....CGGCCAGGAGAAAGTGTCTCGCGGA 888
160 ln.....GlyAlaGlnValProPheArgThrHisSer 170
889 GCGGTACCGGAGAGCGCGCTGGACCCGTCGCGCGCTCCAGTACGTCGAAC 938
171 LeuHisValGlnGlnProPro.....ProGlnGlnProLeuAl 184
939 TCCACGACCGGAGAACCCGCTCGGCGGACCCATCGAGCGCGCGGCTC 988
184 aTyrProLysLeuGlnArgGlnLysLeuGlnAsnAspIleAlaSerProL 201
989 GCGCGCGCTCTCGGTCTCGGCG.....GCCGCGC 1017
201 euProPheProGlnGlyThrHisPheProGlnHis.....Ser 213
1018 GACGACACCCCTCTCTCGGCTCGGCGCAAGACGACGTCGCGGCACTCG 1067
214 GlnSerPheProThrSerSerThrTyrSer...SerSerValGlnGlyG1 229
1068 AAGCGCGCGCGGATCTCTCGGCTCATCAAGACGCTCTCTCGGCTCGGC 1117
229 yGlyGlnGlyAlaHisSerTyrLysSerCysThrAlaProThrAlaGlnP 246
1118 CGGCGCGGATCCGCGGAGCTCAACTTCGTCAGCGCCACCGGACAT 1167
246 roHisAspArgProLeu.....ThrAlaSerSerSer.Le 257
1168 CCCGCTCGACACCTCGGCTCGAGCTGCCCGCGGCTCGGGAGTGC 1217
257 uAlaProGlyGlnArgVal...GlnAsnLeuHisAlaTyrGlnSerGlyA 273
1218 GCGACCGCGGACCGGAACTCTCGCGGCGTCTCGGCTCTCGGCTCGGC 1267
273 rgLeuSerTyrAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 289
1268 GG.....CACCAAGCGCGGCTCTCTCAGCGAGGCGCGCGC 1305
290 GlnGlnAlaLeuGlnSerArgHisHisAlaGlnGlnThrLeuHisTyrG1 306
1306 CCAGGCGCGGAGACCGCGGCTCGATGAGGAGACCGCGCGTGCAGACGG 1355
306 nAsnLeuAlaLysTyrGlnHisTyrGlyGlnGlnGlnGlnGlnGlnGln 323

```

```

1356 GGGCGGCACTGCTCTGTCACCGGCGCGGCGAGGCCCTCGC 1405
1357 GGGCGGCACTGCTCTGTCACCGGCGCGGCGAGGCCCTCGC
323 InPro..AspAlaIaValArgThrProGluGlnIleuTyrGlnPhe 339
1406 GCGGAGCGCGCGGCGCTGCGAGGAGCGGCGAGG.....CGAGCCGGA 1449
339 eProSerSerSerHisSerProAlaArg..... 348
1450 GCTCGGCGCGCGCGGCGAGCTGCGGCGGCGGCGAGCGGCGAGGCTG 1499
349 ..SerValGlyArgSer.....ProSerTyrSerS 358
1500 TCAAGCGACCGGCTGCTGCTCGCGCGCGAGCGCGCGCGCTCTCGAG 1549
358 rThrProSerProLeuMetProAsnIleuGluAsnPheProTyrSerGln 375
1550 GGGCTCGGCGCGCGCTCGCGCGCGGAGCGCGCGCGGCGGCGAGCGG 1599
375 InProLeuSerThrGlyAlaPhePro.....Ala.GlyIleu 387
1600 CACCGCGCGCGCGCGGCGGCGCTGCTGTCAGCGGCGCGAGGCTGCC 1649
387 rAspHisSerHis..PheMetProLeuIleuAsnProSerProThrAspAla 403
1650 AACGTACGGCGATGGCATGAGATTGACGCCCGCGCGCGCGCTTGCAG 1699
404 ThrSerSerValAspThrGlnAlaGlyAsnGlyLysProGluGlnLys 420
1700 ACG...GCCCTGACGCGCGCTCGCGCGCGAGCTGAGCCCTCTCGACG 1746
420 pLysIleuProGluAsnLeuLeuSerAspLeuSerLeuGlnSerLeuThr 437
1747 GCCCTCGCGGAA..... 1759
437 IalauThrIleuGlnIValGluAsnIleSerAsnThrValGln...GlnLeu 452
1760 ..CTGCTCGCGCGCGGAGAACCTCGACGCGAGCTCCACACACACCC 1807
453 LeuLeuSerLysAlaIaValProGlnIleLysGlyLysAsnLeu..V 469
1808 GCGCTCTTCGCGGTGAGGTGCGCTCCACGCGCGCTGAGTACCTCGGG 1857
469 alSerArgThrProGluGlnHisLysSerGlnHisLysSerProGluGly 485
1858 GGTACGCGCGAC.....TectCGCGCGCGAC 1886
486 SerGlyTyrSerAlaGluProAlaGlyThrPro.....LeuSe 498
1887 CCGTCGCGAGATCAGCGCGCGCGCGCGGCTCGTCTCGCGC 1936
498 rGluProProSerSerThrProGln.....SerThrHisAlaGlu 512
1937 GACGCGCGCGCTGCTGCGGCGCGCGCGCGCTCATGACAGCGCTCC 1986
512 rGlnGlnIleAlaAspTyrLeuSerGlySerGluAspProLeuGlnArgSer 528
1987 CGAGGCGCGCGCGAGTGG...TCGCGGTGAGGAGGAGCGAGGAGTGC 2033
529 PheLeuTyrCysAsnGlnAlaArgLys..... 538
2034 TTCGCGACCTCGCGGAGCGAGCGAGCTGCTCGCGCGCGCGGTAGC 2083
539 ....ProAlaArgValAsnSerAsnSerLysAlaLysProGluSerValS 554
2084 GGGCGCGCGCGGCTGCTGCTCGCGCGCGCGAGCGCGCTCTCGACGT 2133
554 er...ThrCysSerVal..ThrSerProAspAspMetSerThrLysSer..A 569
2134 CCGCGAGCTGCTGCGAGAGGCGCGCGCGCGAGCGCGCTCTCT 2183
569 spAspSerPheGlnSerLeuHisGlySerLeuProLeuAspSerPheSer 585
:: ::::: ::::: :::::
2184 CGGACGCTTCGACGCGCGCTAGAGAGCGGAGCTGCGACGCTTCGCG 2233
586 LysPheValAlaGlyIleuArgPoc.....ProArgLeuLeuSe 600
2234 CGG...GTCTGAGAGAGCTGACTTCCAGAGACCGCGCTC..... 2272
600 rAlaLeuAlaGlnIleuAspLeuAlaSerGluIleLeuGlnIleuAla 617
2273 .....GACGTGCTGTCAGGCTGAGAGCGGCGCTGCTGCTGCA 2306
617 IalIleGlyIleuLysAlaAspLysAlaTrpAlaGluAlaProSerLeuVal 633
2307 CACGCGCGAA.....TegAGCGAT...CCGAGACTAGG 2338
634 LysAspSerSerLysProProPheSerLeuGluAsnHisSerAlaCysLe 650
2339 GTGAGACAGTCCCGAGCGCGCTGACGCTC.....CT 2370
650 uAspSerValAlaLysSerAlaTrpProArgProGlyGluProGluAlaL 667
2371 CGAGCGCGTA.....CGGACCTGAGAGAGATCGGCGC 2402
667 eProAspSerLeuGlnIleuAspLysGlyIleuAlaLysAspPheSer 683
2403 CGGACACCTCTCTGAGCTGCTGCGCGAGGCTGCTCGCGATGCGC 2452
684 ProGlyLeuPheGluAspProSerValAlaPheAlaThr..... 696
2453 CGGAGTCCGCTGACCGGACGAGAGGCGCGCGCGCTGCGCGCTGCGC 2502
697 .....ProAspProLysLeuThrThrGlyProLeu..... 706
2503 CAGGCGCGCGCGAGCGCGAGCTGCTGCGCGCGACCTGCGACCGCTCT 2552
707 .....SerPheGlyThrLysPro 712
2553 TCGTCCGCGGCGACAGAGTGCAGTACCGCGCGGAGGAGACCGCGC 2602
713 ThrLeuGlyValProAlaProAspProThrThrAlaAlaPheAsp..... 727
2603 ACGGTAGAGGTGCGCGCTCGCG.....ACCTAGCGCTTCAGCGCGCA 2643
728 .....CysPheProAspThrThrAlaAlaSerAlaAspSerAla 742
2644 ACGCGACCTGCTTCGAC.....GCGCGCGCGAGAGCGCGC 2678
742 snProPheAlaTrpProGluGluAsnLeuGlyAspAlaCys...ProArg 757
2679 CGCGGCTACGCGCGCGCGCGATCGGCGACCGGCTGCGCGCGCGCGC 2728
758 TrpGlyLeuHisProGlyIleuThrLysGlyLeuGlnGlnGlyLys 774
2729 GCGGAGTGTACGCGCGCGAG...GCGAGGCGAGGCGAGGCGCGCGC 2775
774 sAlaSerAspGly..IleSerLysGlyAspThrHisGlnAlaSerAlaCys 790
2776 TCGGAGTGTGCGGTGATCGCGCGCGCTGCGCGAGAGAGACCGAGCGCTGC 2825
791 LeuGlyPheGlnIleuGluAspProProGlyIleuLysValAlaSerLeuPr 807
2826 .....GCGCACAGTCCGCGCGCGCTGCGAGTACAGACAGCCAGC 2866
807 oGlyAsp..PheLysGlnIleuGluValGlyLysValLysGluAlaGly 823
2867 CGCGTGAACCGCGCTCAGCTTCAGAGA.....GCTGGC 2901
824 GlyLeuGlnIleuCysProGluValAlaLysAlaAspArgTrpLeuGluAs 840
2902 .....CTTCGACTCCCTCATGTCGCTGAGACT.....GCGAGAAC 2936
840 pSerArgHisCysSerThrAlaAspPheGlyAspLeuProLeuAsp 857
2937 CGCTGTGACGACAGAGGAGCTGCCCTGCCAGCGAGCTGCTTTCGAC 2986

```







Percent Similarity: 87.755 Percent Identity: 75.510

alignment\_block:

US-09-508-821B-6 x A1893920

Align seg 1/1 to: A1893920 from: 1 to: 963

```

21 SerGlnGlnThrSerArgLeuGlnAsnTyrArgGlnProSerGlnAlaG1 37
1  TCCAGAGAGATCGCCCTGGAGAACTACAGCAGCCGGGTCAGGCTGG 50
37 yLeuSerCysAspArgGlnArgLeuAlaLysAspTyrTyrAsnProG 54
51 GCTAAGCTGTATCGGCAGCGCTGCTGCCAGGACTACTACAGCCCTC 100
54 lnProTyrProSerTyrGlnGlyAlaGlnThrProSerGlyThrAla 70
101 AGCCCTATACAGGCTATGAGGTGGCACTGTACACTTCTGGCAGGCTG 150
71 AlaAlaValAlaAlaAspLysTyrHisArgGlySerLysAlaLeuPro 87
151 GCCACAGAGCTGTCAGCACTGACAGTACAGCAGGAGCAGCAATCC. 192
87 rGlnGlnGlyLeuGlnGlyArgProAlaPheProGlyTyrGlyValGln 104
193 .....CTGCAGGGAGGCGCAGCTTCCCGAGCTAT...GTCAAG 229
104 spSerSerProTyrProGlyArgTyrAlaGlyGlnGlnSerLeuGlnAla 120
230 ACAGCAGCCCTACCCAGGGCGTACTCTGGCGAGGAGGTCTTCAGACC 279
121 TrpGlyAlaProGlnProProProGlnProGlnProGlnProLeuPro 137
280 TGGGGGGGGCCACAGCCAGCCCTCTCAGCCAGCAGCTCTGCCCGGGGC 329
137 yAlaAlaLysTyrAspGlnLeuMetLysTyrThrAlaVal...ProP 153
330 AGTGAGCAAGTATGAGGAGAACCTGTGAGGAGAGACAGCTTGTGCTTC 379
153 rSerArgGlnTyrAlaGlnGlnGlyAlaGlnValProPheArgThrHis 169
380 CAAACAGGCACTACCTGAGCAGGCGCCCGCCAGCTTCTTCGGACTCAC 429
170 SerLeuHisValGlnGlnProProProProGlnProGlnProLeuAla 186
430 AGCTGTCAGTGC...CCACCACACAGCTCAGCAGCCCTGGCTTACCC 476
186 oLysLeuGlnArgGlnLysLeuGlnAsnAspLysLeuSerProLeuPro 203
477 CAACTCCAAAGCAGGAGAAACCCACAGACGACCTTGTCTCCCTGCTCC 526
203 heProGlnGlyThrHisPheProGlnHisSerGlnSerPheProThrSer 219
527 TCCACAG..GGCAGCCACTTCCCGAGCATTCAGATTCCTCCCTACCTCC 575
220 SerThrTyrSerSerSerValGlnGlyGlyGlnGlyAlaHisSerTy 236
576 TCACCTATGCCCCACAGATGTCAGGTTGGTGGGAGGAGGCGCCACTCC 625
236 rLysSerCysThrAlaProThrAlaGlnProHisAspArgProLeuThrA 253
626 ACAGAGCTGCACACACACACTCTGCCAGCTCTCATGATAGGCGGATG 675
253 laSerSerSerLeuAlaProGlyGlnArgValGlnAsnLeuHisAlaTyr 269
676 CCAATGCGAACCTGGGCTCAGGCGNACGGGTCAGAAATCTCACCTTAC 725
270 GlnSerGlyArgLeuSerTyrAspGlnGlnGlnGlnGlnGlnGlnGln 286
726 CAGCTGGCGGCTTGGCTAC..... 746
286 nGlnGlnGlnGlnGlnAlaLeuGlnSerArgHisHisAla..GlnGln 302

```

```

747 .GAGCAGCAGCAGCAGCAGCTTCAAGGGCGCTCACCACACCCAGGAACA 795
303 LeuHisTyrGln.AsnLeuAlaLysTyrGlnHisTyrGlnGlnGlyG 319
796 CTTCACTACAGAAACCTGGCCAGTACCACTATGGACACAGGATAC 845
319 lngLysTyrCysGlnProAspAlaAlaValArgThrProGlnGlnTyr 335
846 AGGCTACTCTCCACCGGACACAGCTGTCCAGGACTCCACACAGGATAC 895
336 GlnThrPheSerProSerSerHisSerProAlaArgSerValGly.A 352
896 .GAGACTTTCAGCCCGGACTCCAGCCATCCCTGTCAGCT...CTGGGTTC 942
352 rgSerProSerTyrSerSerThr 359
943 GT...CCCTTCTATAGCTCCACC 962

```

seq\_name: gb\_est2:BE869510

seq\_documentation\_block: 865 bp mRNA EST 20-OCT-2000  
LOCUS BE869510  
DEFINITION 601446178F1 NIH\_MGC\_65 Homo sapiens cDNA clone IMAGE:3850146 5',  
mRNA sequence.

ACCESSION BE869510  
VERSION BE869510.1 GI:10318286  
KEYWORDS EST.

SOURCE

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 865)

AUTHORS NTH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgaabs-re@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1AM9568 row: 1 column: 19

High quality sequence stop: 662.

Location/Qualifiers

FEATURES

source

```

1..865
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3850146"
/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotL;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

```

BASE COUNT 184 a 294 c 242 g 145 t  
ORIGIN

alignment\_scores:

Quality: 1088.50 Length: 260  
Ratio: 4.612 Gaps: 4  
Percent Similarity: 90.769 Percent Identity: 85.385

alignment\_block:

US-09-508-821B-6 x BE869510

Align seg 1/1 to: BE869510 from: 1 to: 865

1499 LysMetGlnGlnLeuGly...ProAlaSerGlnProGlnGlnArgPr 1514

```

|||||
1 AAGATGAGAGAGCTGGCCCTGCTCCCGACGCCCCGCTGAGGCGCAGGCC 50
1514 OCYGLINPROGLINThrArgLysGlnProGlyHisThrSntYs 1531
|||||
51 CTGGCAGCCCGACAGGCGACAGAAACAGGCGCAGCCACCACTACA 100
1531 erSerTySerLysArgLysThrLeuThrArgLysAlaLysAsnThr 1547
|||||
101 GCAGCTATTCACAGCGGAGCCCTCACTCGGGGCGCGGAGAAACAC 150
1548 ThrSerSerProCysLysGlyArgAlaLysArgArgArgGlnGlnGln 1564
|||||
151 ACCTCTTCCCTGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 200
1564 LLeuProLeuAspProAlaGluProGluLysLeuLysTyrlleSers 1581
201 GCTGCCCCCTGATCCCGAGAGCCCTGAATCCCGCTCAAGATATTC 250
1581 erCysLysArgLeuArgSerSerSerArgThrProAlaSerProPhe 1597
|||||
251 CTGGCAGGCGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 300
1598 ValArgValGluLysArgAspAlaPheThrThrLecysThrValAla 1614
301 GTGCGGGGTGAGAGAGCGAGCGCTTACACACATGACAGCTTGTACA 350
1614 nSerProGlyAspAlaProLysProHisArgLysProSerSerSer 1631
351 CTCCCTCGAGATGCGCCCAAGCCCAAGAGAGAGAGAGAGAGAGAG 400
1631 erSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 1647
401 CTTCTTCCCTGATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 450
1648 LeuAlaThrLeuProGlyLysThrLeuGlnProArgProSerLeu 1664
451 CTGGCCACAGTCCCTGAGAGCTCCATCCCTGAGAGCCCTGAGCC 500
1664 OlauSerSerTherMetHisLeuGlyProValValSerLysAlaLeu 1681
501 CTTCTTCCCTGAGATGAGTGGGCGCTGCTGCTGCTGCTGCTGCT 550
1681 hSerCysLeuValCysCysLeuGlnAsnProAlaAsnPhelys 1697
551 CTTCTTCCCTGATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
1697 PleuGlyAspLeuCysGlyProTyThrProGluHisCysLeuPro 1714
601 CTTCTTCCCTGATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 650
1714 yLysProLysLeuLysGlnLysValArgProGlnGlyThrCysGln 1730
651 AGAAGCAAACTCCAGAGAGAGTGGCGGAGAGAGAGAGAGAGAG 700
1731 AlaSerLeuProLeuGluArgThrLeuLysGlyProGluLysAla 1747
701 GCTCTGCTGCGCTGAGAGAGACACAAAGTGTCCGAGTGTGGAAT 750
1747 laAlaThrAlaGlyLysProPro 1754
751 CCGAGACAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 774
seq_name: gb_est2:B6828494
seq_documentation_block:
LOCUS B6828494 724 bp mRNA EST 22-MAY-2001
DEFINITION 602751486F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4904229 5',
mRNA sequence.
ACCESSION B6828494
VERSION B6828494.1 GI:14176081
KEYWORDS EST.
SOURCE human.

```

```

ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 724)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILMIL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILMIL at:
http://image.llnl.gov
Plate: LHC1803 row: 1 column: 22
High quality sequence stop: 724.
FEATURES
source location/Qualifiers
1..724
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4904229"
/clone_id="NIH_MGC_17"
/tissue_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: muscle; Vector: pOT8; Site:1: EcoRI;
Site:2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAAGAG(C). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies). "
BASE COUNT 161 a 246 c 217 g 100 t
ORIGIN
alignment_scores:
Quality: 996.50 Length: 211
Ratio: 4.982 Gaps: 5
Percent Similarity: 94.787 Percent Identity: 93.365
alignment_block:
US-09-508-821b-6 x B6828494
Align seg 1/1 to: B6828494 from: 1 to: 724
1 MetGlnSerPheArgGluArgCysGlyPheHisGlyLysGlnGlnSntY 17
|||||
101 ATGCAAGTCTTTTCGAAAAGGTGTGCTTCAGAGGCAACACAGACTA 150
17 rgGlnThrSerGlnGluThrSerArgLeuGlnGlnSntYrArgGln. Pr 33
|||||
151 CCGAGAGACCTCGAGAGAACTCAACCCCTAGAGATTCAGTGGCAAGCC 200
33 oSerGlnAlaGlyLeuSerCysAspArgGlnArgLeuAlaLysAsp 50
|||||
201 GAGTCAGGCGCGGCTAAGCTGAGACCGGAGCGGCTGCTGCCCAAGACT 250
50 YTYrAsnProGlnProTyThrProSerTyrlleGlyAlaGlyThrPro 66
|||||
251 ATTATTAACCGCAGCCTTACCGGAGCTATGAGGAGGCGGCGGAGCC 300
67 SerGlyThrAlaAlaValAlaAlaAspLysTYrHisArgLysSerly 83
|||||
301 TCGGCACTGAGAGCGGCTGAGCGCGGAGACATACACAGGAGAGCA 350
83 salaleuProThrGlnGlnGlyLeuGlnGlnGlyArgProAlaPhePro 100
|||||
351 GGCCCTGGCCCAAGAGAGGCTGAGGAGGAGAGGCGGCTTCCCGGCT 400
100 YrGlyValGlnAspSerSerProTyThrProGlyArgTyAlaGlyGln 116

```



```

|||||
401 ACGGGTCCAGGACAGAGCCCTTACCCAGGCGCTATGCTGGTGAGAG 450
|||||
117 SerLeuGlnAlaTrpGlyAlaProGlnProProProGlnProGlnPro 133
|||||
451 AGCCTTCAGGCTTGGGGGCCCCACAGCCACCAACCCACAGCCGAGCC 500
|||||
133 cLeuProAlaGlyValAlaLysTyrAspGlnLysLeuMetLysLysThra 150
|||||
501 ACTACCTGCAGGGGTGGCAAGTATGATGAGAACTTGATGAAAAAGACAG 550
|||||
150 laValProProSerArgGlnTyrAlaGlnGlnAlaGlnValProPhe 166
|||||
551 CAGTCCCCCCCCACAGCAGTATCAGACAGAGGGGCCGCCAGTGCCCTTT 600
|||||
167 ArgThrHisSerLeuHisValGlnGlnProProProGlnGlnPro 182
|||||
601 CGGACTCAGCTCCCTTGCACCGTCCAGCAGACACCCGCCGCCAGCAGACC 650
|||||
183 LeuAlaTyrProLysLeuGlnArgGlnLysLeuGlnAsnAspIleAla 199
|||||
651 TGG...CATACCCCAAGCTCCAAAGGCAAGAGCTGAGAAAGACATTTG 697
|||||
199 exProLeuProPheProGlnGlyThr 207
|||||
698 CCGCTCTGACCTTCCGCCAGGGTACC 723

```

seq\_name: gb\_est2:BF307287

```

seq_documentation_block:
LOCUS      BF307287      1012 bp      mRNA      21-NOV-2000
DEFINITION 601891563F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4137399 5',
            mRNA sequence.
ACCESSION  BF307287
VERSION    BF307287.1 GI:11254402
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 1012)
            NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: AFCC
            cDNA library Preparation: Ling Hong/Rubin Laboratory
            cDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM1044 row: m column: 16
            High quality sequence stop: 736.

```

```

FEATURES             location/qualifiers
     source           1..1012
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4137399"
                     /clone_lib="NIH_MGC_17"
                     /tissue_type="inadomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: muscle; Vector: pORF7; Site_1: EcoRI;
                     Site_2: XhoI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GCCACGAG(G). Size-selected >500bp
                     for average insert size 1.8kb. Library constructed by
                     Ling Hong in the Laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT          260 a      307 c      295 g      148 t      2 others
ORIGIN
|||||

```

```

alignment_scores:
    Quality: 989.50      Length: 292
    Ratio: 4.039        Gaps: 10
    Percent Similarity: 83.904      Percent Identity: 74.658

alignment_block:
US-09-508-821B-6 x BF307287 ..
Align seg 1/1 to: BF307287 from: 1 to: 1012

1 MetGlnSerPheArgGlnLysCysGlyPheHisGlyLysGlnGlnAsnTyr 17
2 AFGCAGTCTTTTCGAGAAAGGTGTGGTTTCCATGCAACACAGAACTA 151
3 rGlnGlnThrSerGlnGlnThrSerArgLeuGlnAsnTyrArgGlnPro 34
4 CCAGCAGACCTCGCAGGAACATCATCGCTAGAGAAATTACAGCAGCCGA 201
5 erGlnAlaGlyLeuSerCysAspArgGlnArgLeuAlaLysAspTyr 50
6 GTCAGGCGGGCTTAAGCTCGACCGCAGCGGCTGCTGCCAAGGACTAT 251
7 TyrAsnProGlnProTyrProSerTyrGlnGlyAlaGlyThrProSe 67
8 TATACCGCGAGCTTACCGAGCTATGAGGTGGCTGGCAGGCCCTC 301
9 rGlyThrAlaAlaValAlaLysPlyTyrHisArgGlySerLysA 84
10 TGGCAGCTGCAGCGGGTGGCGCGCACAAAGTACCACCGAGCGCAGG 351
11 laLeuProThrGlnGlnGlyLeuGlnArgProAlaPheProGlyTyr 100
12 CCGTGGCCGCACAGCAGAGCGCTGCGAGGGAGCGCGCTTCCCTGGCTAC 401
13 GlyValGlnAspSerSerProTyrProGlyArgTyrAlaGlyGluSe 117
14 GCGCTGCCAGCAGCAGCGCCCTACCCAGCGCGCTATGCTGGTGAGGAG 451
15 rLeuGlnAlaTrpGlyAlaProGlnProProProProGlnProGlnPro 134
16 CCGTTCAGGCTGGGGGGGGCCCCACAGNCAACAACCCACAGCGCGCAGCA 500
17 euProAlaGlyValAlaLysTyrAspGlnAsnMetLysLysThra 150
18 TACCTGCGAGGGTGGCGCAGTATGTGAGAACTTGTGAAAGACAGCA 550
19 ValProProSerArgGlnTyrAlaGlnGlnGlyAlaGlnValProPhe 167
20 GTGCCCCCAGCAGCAGGCTATGCGAGCAGGCGGCCAGGTTGCCCTTCG 599
21 qThrHisSerLeuHisValGlnGlnProProProProGlnGlnProLeu 184
22 GACTCATCCCTCCAGCTCCAGCAGCAACAGCAGCGCGCCCTGT 649
23 laTyrProLysLeuGlnArgGlnLysLeuGlnAsnAspIleAlaSer 200
24 GATACCCAGCTCCCAAGGCGAGAGCTTGCAGACAGGACATNGGCTCC 699
25 oLeuProPheProGlnGlyThrHisPheProGlnHisSerGlnSerPhe 216
26 TCTTGGCTTCCCA...GGGTACCACTTTCCTCAGGATTCAGTCTT... 743
27 ProThrSerSerThrTyrSerSerValGlnGlyGlyGlnGlyAl 233
28 TCCAGACTCTTCCATATTTCTCTCTCTCAGTGGGGGAGCGGG... 788
29 ahSerTyrLysSerCysThrAlaProThrAlaGlnProHisAspArg 250
30 CGCATCTATAAGAGT.....TGACAGACGAATGGCAAC 822
31 roLeuThrAlaSerSerSerLeu...AlaProGlyGlnArgValGln 265

```



Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et3-RC0-CT0379-290  
100-032-clit3-2000-01-29at4-1)  
Seq primer: puc 18 forward  
High quality sequence start: 11  
High quality sequence stop: 566.

## FEATURES

Location/Qualifiers  
1..566  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="CT0379"  
/dev\_stage="Adult"  
/note="Organ: colon; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 99 a 211 c 149 g 107 t

## ORIGIN

alignment\_scores:  
Quality: 819.00 Length: 160  
Ratio: 5.184 Gaps: 1  
Percent Similarity: 98.750 Percent Identity: 98.750

## alignment\_block:

US-09-508-821B-6 x AW860154

Align seg 1/1 to: AW860154 from: 1 to: 566

1597 PheValArgValGluLysArgAspAlaPheThrThrIleCysThrValva 1613  
|||||  
13 TTCGTGGGGTGGAGAGCAGAGCGGTTCACCCACCATATGACATGTTGT 62  
|||||  
1613 LasSer.ProGlyAspAlaProLysProHisArgLysProSerSerSer 1629  
|||||  
63 CAATCCCGCTGGAGATGCCGCCAGCCAGCCAGAGAGCTTCCTCCTCT 112  
|||||  
1630 AlaSerSerSerSerSerSerSerPheSerLeuAspAlaIleAl 1646  
|||||  
113 GCCTCCTCTTCCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 162  
|||||  
1646 aSerLeuAlaThrLeuProGlyGlySerIleLeuGlnProArgProSerL 1663  
|||||  
163 CTCCCTGGCCACACTCCCTGGAGGCTCCATCTGACGCCGGCCCTCT 212  
|||||  
1663 euProLeuSerSerThrMethisLeuGlyProValValSerLysAlaLeu 1679  
|||||  
213 TCCCTCCTCTCCAGATGACATGGGGCTGTGTGTTCAGAGCCCTG 262  
|||||  
1680 SerThrSerCysLeuValCysCysLeuGlnAsnProAlaAsnPhely 1696  
|||||  
263 AGTACCTCTTGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 312  
|||||  
1696 sAspLeuGlyAspLeuGlyProGlyProGlyProGlyProGlyProGly 1713  
|||||  
313 GGACCTTGGGACCTCTGTGGCCCTACTACCTTGACACACTGCTCCCA 362  
|||||  
1713 ystLysLysProLysLeuLysGluLysValArgProGluGlyThrCysGlu 1729  
|||||  
363 AAAAGAGCCAAACACTCAAGGAGAGGTGGCCAGCAGGACCTGTGAG 412  
|||||  
1730 GluAlaSerLeuProLeuGluArgThrLeuLysGlyProGluCysAlaAl 1746  
|||||

413 GAGGCTCGCTGCCGCTTGAGAGACACTCAAGAGTCCGAGTGTGCAGC 462

1746 aAlaAlaThrAlaGlyLysProProArg 1755

|||||  
463 TGCGCGCACTGCCGGAAGCCGCCAGC 490

seq\_name: gb\_est1:AW860199

seq\_documentation\_block: 566 bp mRNA EST 19-MAY-2000  
LOCUS AW860199  
DEFINITION RCO-CT0379-060200-032-cl1 CT0379 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AW860199  
VERSION AW860199.1 GI:7955892  
KEYWORDS EST.  
SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

1. (bases 1 to 566)

## AUTHORS

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordino, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,  
Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.

## TITLE

Shotgun sequencing of the human transcriptome with ORF expressed

## JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

## MEDLINE

20202663

## COMMENT

Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et3-RC0-CT0379-060  
200-032-clit3-2000-02-06at4-1)  
Seq primer: puc 18 forward  
High quality sequence start: 11  
High quality sequence stop: 566.

BASE COUNT 99 a 211 c 149 g 107 t

ORIGIN

## FEATURES

## source

1..566  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="CT0379"  
/dev\_stage="Adult"  
/note="Organ: colon; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 99 a 211 c 149 g 107 t

ORIGIN

alignment\_scores:  
Quality: 819.00 Length: 160  
Ratio: 5.184 Gaps: 1  
Percent Similarity: 98.750 Percent Identity: 98.750

## alignment\_block:

US-09-508-821B-6 x AW860199

Align seg 1/1 to: AW860199 from: 1 to: 566

1597 PheValArgValGluLysArgAspAlaPheThrThrIleCysThrValva 1613

|||||

seq\_name: gb\_gss:AZ642061

ACCESSION	AZ642061
VERSION	AZ642061.1
	GI:11768322

**KEYWORDS** GSS.  
**SOURCE** house mouse.

ORGANISM

house mouse.

[illegible]

## REFERENCE AUTHORS

**TITLE**

JOURNAL  
COMMENT

University of Utah Genomic Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801.581.5500

Tel: 801 585 5606  
Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000      Std Error: 0.00

Plate: 0504 Row: I Column: 24

```

    def primer : String = "CTGTTTGGAGGCGAATTCACCACTC"
    class plasmid ends

```

High quality sequence stop: 501.

## FEATURES

Location/qualifiers

**Source**

1. 501  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UG6C1M0504I24"  
/clone\_1kb="mouse 10kb plasmid UG6C1M library"  
/sex="Male"  
/lab\_host="E. coli strain XL10-Gold,  $\pi$ -resistant, F-"  
/note="Vector: pMD24v; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse Services/documents/dnares/). The DNA  
was hydroxymethylated, sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pMD2 (g147311419b) (AF129072.1), a copy-number  
inducible derivative of plasmid RL. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT  
ORIGIN :

```
alignment_scores: 165
    quality: 807.00
    ratio: 4.581
Percent Similarity: 98.182
Percent Identity: 93.333
```

```
alignment_block;
```

Align: seg 1/1 to: AZ642061 from: 1 to: 501

350 ValGlyArgSerProSerTyrSerSerThrProSerProLeuMetProAs 366

3 GTGGGTGCTCCCTTCTTATAGCTCCACCCCGTCACCACTGATGCCAA 52

366 110001468BronturaSerj1nG1nPro1auSer1hrg1va1aPhep 383

100

383 TOWALY IRENE ASPHISSEHNSFHEMELI VOLUCULUMBN 1000

103 CCACAGGCATCACAGACCCACAGCCACTTATGCCCCCGCTAACCCGTC 157

400 ProThrAspAlaThrSerSerValAspThrGlnAlaGlyAsnGlySer 410

153 CCAACAGATGCTGCCAGCTCTGTGACCCCCCAGGCCGGCACTGCAAGCC 20

416 ofLengthInFvsAspLvsLysLeuProGluAsnIleuLeuSerAspLysLeuG 43

25

bioRxiv preprint doi: <https://doi.org/10.1101/000000>; this version posted January 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[illegible][illegible]

450 GINGINLEULEUSERLYSALAALVAALPROGIMLYSLSYVALDY 40

303 CAGCAGCTTGTCTGTCCAAGCTTACCATTGCCACAGAGAAGGGGTCAA 35

466 sasLeuValSerArgThrProGluGlnHisLysSerGlnHisCysSerP 48

353 GAACCTCGTGTCCAGGACTCCAGAGCAGCACAAAGCCAGCACTGTACCC 40

483 rogluGlySerGlyTyrSerAlaGluProAlaGlyThrProLeuSerGlu 499  
 |||||  
 403 CTGAGGAGCGGCTTCTGAGTGAGCGAGCGGACACCGCTGTCTGAG 452  
 |||||  
 500 ProProSerSerThrProGlnSerThrHisAlaGluProGlnGlu 514  
 |||||  
 453 CGCGGAGCAGCAGCAGCAGTCCACCATGCTGAGCCACAGGAC 497  
 |||||

seq\_name: gb\_est1:AW860153

seq\_documentation\_block: 558 bp mRNA EST 19-MAY-2000  
 LOCUS AW860153  
 DEFINITION RCO-CT0379-290100-032-b08 CT0379 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION AW860153  
 VERSION AW860153.1 GI:7955846  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 558)  
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
 Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
 ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
 Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed

JOURNAL sequence tags  
 MEDLINE Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 COMMENT 20202663

Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-RCO-CT0379-290  
 100-032-b08&t3=2000-01-29&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 38

High quality sequence stop: 558.

FEATURES Location/Qualifiers

source

1..558

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="CT0379"

/dev\_stage="Adult"

/note="Organ: colon; Vector: puc18; Site: 1: SmaI; Site 2:

SmaI; A mini-library was made by cloning products derived

from ORSTES PCR (U.S. Letters Patent application No. 196

,716 - Ludwig Institute for Cancer Research) profiles

into the pUC 18 vector. Reverse transcription of tissue

mRNA and cDNA amplification were performed under low

stringency conditions."

BASE COUNT 98 a 207 c 147 g 106 t

ORIGIN

alignment\_scores:

Quality: 794.00

Ratio: 4.994

Percent Similarity: 98.758

Length: 161

Gaps: 1

Percent Identity: 97.516

alignment\_block:

US-09-508-821B-6 x AW860153

Align seg 1/1 to: AW860153 from: 1 to: 558

1506 PropheValArgValGluLysArgAspAlaPheThrThrIleCysThrVa 1612  
 |||||  
 3 CCTCTGCTGGGTGGAGAGGA.GACGGCTTACACCAATGACACTGT 51  
 |||||  
 1612 lValAsnSerProGly AspAlaProLysProHisArgLysProSerSer 1628  
 |||||  
 52 TGTCAACTCCCTGTGAGATGCGCCCAAGCCCAAGAGAGCCCTCTCC 101  
 |||||  
 1629 SerAlaSerSerSerSerSerSerSerSerPheSerLeuAspAlaAa 1645  
 |||||  
 102 TCTGCTCTCTCTCTCATCTCTCTCTCTCTCTCTCTCTCTCTCT 151  
 |||||  
 1645 yAlaSerLeuAlaThrLeuProGlyGlySerIleLeuGlnProArgPro 1662  
 |||||  
 152 GGCCTCCCTGGCCACACTCCCTGGAGGCTCCATCTCGACGCCGCGC 201  
 |||||  
 1662 erLeuProLeuSerSerThrMetHisLeuGlyProValValSerLysAla 1678  
 |||||  
 202 CTTGGCCCTCTCTCCACAGTACCTTGGGGCTGTGTGTTTCCAAAG 251  
 |||||  
 1679 leuSerThrSerCysLeuValCysCysLeuCysGlnAsnProAlaAsnPh 1695  
 |||||  
 252 CTGAGTACCTCTTGGCTTGTGTGCTCTGCTGCTGCTGCTGCTGCT 301  
 |||||  
 1695 eLysAspLeuGlyAspLeuCysGlyProTyrTyrProGluHisCysLeuP 1712  
 |||||  
 302 CAAGGACTTGGGGACCTCTGTGCTGCTGCTGCTGCTGCTGCTGCT 351  
 |||||  
 1712 roLysLysLysProLysLeuLysGlyLysValArgProGluGlyThrCys 1728  
 |||||  
 352 CCAGAGAGAGAGGCAAACTCAAGGAGAGAGTGGCGCCAGAGGACCT 401  
 |||||  
 1729 GluGluAlaSerLeuProLeuGluArgThrLeuLysGlyProGluCysAl 1745  
 |||||  
 402 GAGGAGGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 451  
 |||||  
 1745 alaAlaAlaThrAlaGlyLysProProArg 1755  
 |||||  
 452 AGCTGCCGACACTGCCGGAAGAGGCCCCCAGG 482

seq\_name: gb\_est1:AW862383

seq\_documentation\_block:

LOCUS AW862383

DEFINITION RCO-CT0379-060200-032-b08 CT0379 Homo sapiens cDNA, mRNA sequence.

ACCESSION AW862383

VERSION AW862383.1 GI:7958081

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 558)

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,

Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,

Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?l=et2-RCO-CT0379-060  
200-032-b08et3-2000-02-06et4-1)  
Seq primer: puc 18 forward  
Seq quality sequence start: 38  
High quality sequence stop: 558.  
Location/Qualifiers

FEATURES

source  
1. 558  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="CT0379"  
/dev\_stage="Adult"  
/note="Organ: colon; Vector: puc18; Site\_1: Sma1; Site\_2:  
Sma1; A mini-library was made by cloning products derived  
from ORESTES PCR (U.S. Letters Patent application No. 196  
716 - Ludwig Institute for Cancer Research) profiles  
into the pUC 18 vector. Reverse transcription of tissue  
mRNA and cDNA amplification were performed under low  
stringency conditions."

BASE COUNT 98 a 207 c 147 g 106 t  
ORIGIN

alignment\_scores:  
Quality: 794.00 Length: 161  
Ratio: 4.994 Gaps: 1  
Percent Similarity: 98.758 Percent Identity: 97.516

alignment\_block:  
US-09-508-821B-6 x AN862383 ..

Align seg 1/1 to: AN862383 from: 1 to: 558

1596 prophevalaryglulysargaspalaphethrthrlleCysThrva 1612  
|||||.....  
3 CCTTCGTGGGGGTGGAGAGGA. GAGCCTTACCAACATGCACTG 51

1612 ValanserProGly. AspAlaProlyProHisArglyProSerSer 1628  
|||||.....  
52 TGTCAACTCCCTCTGTGAGATGGCCCAAGCCCAAGAGCCTTCTCC 101

1629 SerAlaSerSerSerSerSerSerSerSerSerSerSerSerSer 1645  
|||||.....  
102 TGTGCT 151

1645 ValaserleuAlaThrleuProGlySerleuLeuGlnProArgPro 1662  
|||||.....  
152 GGCCTCCCTGGCCACACTCCCTGGAGGCTCCATCTGCAAGCCGCC 201

1662 extleuProleuSerSerThrMetHisleuGlyProValSerlyAla 1678  
|||||.....  
202 CCTTGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 251

1679 leuSerThrSerCysleuValCysCysleuGlnAsnProAlaSerPh 1695  
|||||.....  
252 CTGAGTACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 301

1695 elysaspIleuGlyaspIleuGlyProGlyThrProGlyHisCysLeu 1712  
|||||.....  
302 CAAGGACCTTGGGACCTCTGTGGCCCTACACCTGAGACACTGCTCC 351

1712 rGlylySerProlySerleuValGlyValAlaProGlyGlyThrCys 1728  
|||||.....  
352 CCAAAAGAGCCAAACTCAAGAGAAAGAGAGAGAGAGAGAGAGAGAG 401

1729 GlnGlnAlaSerleuProleuGlnArgThrleuGlyGlyProGlyAla 1745  
|||||.....  
402 GAGGAGGCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 451

1745 AlaAlaAlaThrAlaGlylySerProArg 1755  
|||||.....  
452 AGTCGCGCCACGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 482

seq\_name: gb\_est1:AA052617

seq documentation block:  
LOCUS AA052617 611 bp mRNA EST 13-SEP-1996  
DEFINITION mc85a07.11 Soares mouse embryo NM0613.5 14.5 Mus musculus cDNA  
clone IMAGE:355284 5' similar to gb:029801 Mouse mRNA for unknown  
product, complete cds (MOUSE);, mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 611)  
REFERENCE  
AUTHORS

Marr, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucata, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Thelning, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.

Mouse NIH Mouse EST Project  
The Mouse NIH Mouse EST Project  
Unpublished (1996)  
Contact: Maria M/Mouse EST Project  
WashU-HMNI Mouse EST Project  
Washington University School of Medicine  
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LML; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:227084

Trace considered overall poor quality  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 1.  
Location/Qualifiers

1. 611  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:355284"  
/clone\_lib="Soares mouse embryo NM0613.5 14.5"  
/sex="unknown"  
/tissue\_type="embryo"  
/dev\_stage="13.5-14.5dpc total fetus"  
/lab\_host="DH10B"

/note="Vector: pT73D-pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5',  
TGTACCAATCTGAGTGGGAGGCGGCGGGAATTTTCTTTTCTTTTCTTTT  
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2  
14.5dpc embryos (total RNA provided by Minoru Ko, Wayne  
State Univ., from 2 ]: double-stranded cDNA was ligated to  
Eco RI adaptors (Pharmacia), digested with Not I and  
cloned into the Not I and Eco RI sites of the modified  
pT733 vector. Library went through one round of  
normalization, and was constructed by Banto Soares and  
M.Fatima Bonaldo."

BASE COUNT 139 a 204 c 165 g 103 t  
ORIGIN

alignment\_scores:  
Quality: 793.50 Length: 219  
Ratio: 4.312 Gaps: 7  
Percent Similarity: 84.018 Percent Identity: 72.146

alignment\_block:  
US-09-508-821B-6 x AA052617 ..

Align seg 1/1 to: AA052617 from: 1 to: 611

22 GlnGlnThrSerArgleuGlnuSerArgGlnProSerGlnAlaGlyle 38  
|||||.....

2 CAGGAGACATCGCGCTGGAGAACTACAGCAGCGCGGTACAGCGTGGGCT 51  
 38 userCysAspArgGlnAlaLeuAlaLysAspTyrTyrAsnProGlnP 55  
 52 AAGCTGTGTATCGGACGCGCTGCTGGCCAGAGACTACTACAGCCCTCAGC 101  
 55 rOTyProSerTyrGluGlyAlaGlyThrProSerGlyThrAlaAla 71  
 102 CTTATACAGGCTATGAGGTGGCTGTGTACACTTCTGGCAGCGTGGCC 151  
 72 AlaValAlaAlaAspLysTyrHisArgGlySerLysAlaLeuProThrG 88  
 152 ACAGCAGGTGAGACAGTACACCGAGGAGCAATCC..... 190  
 88 nGlnGlyLeuGlnGlyArgProAlaPheProGlyTyrGlyValGlnAsp 105  
 191 .....CTGCAGGGGAGGCGAGCTTCCCACTAT.....GTCAAGACA 230  
 105 exSerProTyrProGlyArgTyrAlaGlyGluGlnSerLeuGlnAlaTyr 121  
 231 GCAGCCCTACATAGGCGCTACTCTGCGAGGAGAGTCTTCAGACCTGG 280  
 122 GlyAlaProGlnProProProGlnProGlnProGlnProAlaGlyVa 138  
 281 GGGGGCCACAGCCACCGCTCTCAGCCACAGCTCTGCCCGGGGAGT 330  
 138 lAlaLysTyrAspGluAsnLeuMetLysThrAlaValPro...ProS 154  
 331 GAGCAGTATGAGGAGAACTGATGAGNAGACAGTGTGCTCTTCCAA 380  
 154 erArgGlnTyrAlaGluGlnGlyAlaGlnValProPheArgThrHisSer 170  
 381 ACAGGAGTACCTGAGCAGGCGCCAGCTTCCCTCCGAGTCCACAGC 430  
 171 LeuHisValGlnGlnProProProGlnGln...ProLeuAlaTyrPro 186  
 431 CTGCATGTC.....CCACACACAGCCTCAGCACCCCTGGCTTACGA 474  
 187 LysLeuGlnArgGlnLysLeuGlnAsnAspIleAlaSerProLeuProPh 203  
 475 AATTCCAAAGGCAAAACAGAGAGGACCTGGCCCTCCCTGTGCGCTT 524  
 203 eProGlnGlyThrHisPheProGlnHisSerGlnSerPhe..... 216  
 525 CGCGCAGGAGCAGCACTTTCCAGCAGTCCGAGTCTTCTCAGCTCA 574  
 217 .....ProThrSerSerThrTyrSerSerValGlnGlyGly 230  
 575 ACTWATGCCCAACATGTTCA.....GGTGGTGGC 603  
 231 GlnGly 232  
 604 AAGGG 609

seq\_name: gb\_est2:BF304773

seq\_documentation\_block: 832 bp mRNA 21-NOV-2000  
 LOCUS BF304773 832 bp mRNA EST  
 DEFINITION 601888356F1 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:4122469 5',  
 mRNA sequence.

ACCESSION BF304773.1 GI:11251671

VERSION BF304773

KEYWORDS EST

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 832)

AUTHORS NIH-National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)

JOURNAL Contact: Robert Strausberg, Ph.D.

COMMENT Email: cgabbs@mail.nih.gov

Tissue Procurement: ATCC  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 cDNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov](http://image.llnl.gov)  
 Plate: LLC1005 row: 0 column: 14  
 High quality sequence stop: 619.

# FEATURES

## source

1..832  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4122469"  
 /clone\_lib="NIH\_MGC\_17"  
 /tissue\_type="rhodomyosarcoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: muscle; Vector: pOTB7; Site\_1: EcoRI;  
 Site\_2: XhoI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCACGAG(S). Size-selected >500bp  
 for average insert size 1.8kb. Library constructed by  
 Ling Hong in the laboratory of Gerald M. Rubin (University  
 of California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 195 a 256 c 247 g 134 t  
 ORIGIN

## alignment\_scores

Quality: 791.00 Length: 253  
 Ratio: 3.821 Gaps: 12  
 Percent Similarity: 81.818 Percent Identity: 73.123

## alignment\_block

US-09-508-821B-6 x BF304773 ..

Align seg 1/1 to: BF304773 from: 1 to: 832

1 MetGlnSerPheArgGluArgCysGlyPheHisGlyLysGlnGlnAsnTy 17  
 101 ATGCAGTCTTTTCGAGAAAGGTGGTTCATCGCAACACACAGAACTA 150  
 17 rGlnGlnThrSerGlnGlnThrSerArgLeuGluAsnTyrArgGlnProS 34  
 151 CCAGCAGACCTCGCAGGAACATCAGCCTAGAGAAATTACAGGCAGCGCA 200  
 34 erGlnAlaGlyLeuSerCysAspArgGlnArgLeuAlaLysAspTyr 50  
 201 GTCAGCGCGGCTAAGCTGCGACCGCGAGCGCTGCTGCCAAGGACTAT 250  
 51 TyrAsnProGlnProTyrProSerTyrGlnGlyValAlaGlyThrProse 67  
 251 TATACCGCGAGCCTTACCGAGCTATGAGGTGGCGCTGGCAGCCCTC 300  
 67 rGlyThrAlaAlaAlaValAlaAlaAspLysTyrHisArgGlySerLysA 84  
 301 TGGCAGTCAGCGCGGTGGCGCGACACAGTACCACCGAGGCGACAGG 350  
 84 laLeuProThrGlnGlnGlyLeuGlnGlyArgProAlaPheProGlyTy 100  
 351 CCCTGCCACACACAGCAAGGCTCGAGGGAGGCGCTTCCCTGGGCTA 400  
 100 r.GlyValGlnAspSerSerProTyrProGlyArgTyrAlaGlyGluGlu 116  
 401 CTGGCTCCAGACAGACAGCCCTTACCCAGT.GGTATGCTGGTGAGGAG 449  
 117 SerLeuGlnAlaTyrGlyAlaProGlnProProProProGlnProGlnPr 133  
 450 AGCTTCAGGCTT...GGGGGCCCAAGCCACACACCCATAGCGCAGCAC 496  
 133 oLeuProAlaGlyValAlaLysTyrAspGluAsnLeuMetLysLysThra 150  
 497 TACCTGAGGGGT...GCCAAGTATGATGAGAACTTGTGAAAAAGACAG 543





cc: name: ab est1:22655437

ENTER THE WECHTL-HHMT MOUSE EST PROJECT

alignment block:

1640

1640

68 CAGGAGCGCTCCTCCCTCATCCTCGTCTGTCT 117

000

*[The following section contains numerous small, illegible fragments of text, likely bleed-through from another page.]*

118 CCTTGGATGCAGCCGGGGCTCCCTGGCCACACTCCCTGGAGGCTCCATC 167

	LeuGlnProArgProSerLeuProLeuSerIleIleMechTrsPheuYll	1657
	LeuGlnProArgProSerLeuProLeuSerIleIleMechTrsPheuYll	1693

217

1673 CysValSerIleValAlaLeuSerThrSerCysIleuValCysCysLeuCysG 1690

[illegible]

1690 InAsnProAlaAsnPheLysAspLeuGlyAspLeuCysGlyProTyrTyr 1706

268 AAAACCGGCCAACCTTCAAGGACCTTGGGACCTCTGTGGCCCTACAC 317

1773

318 C C T G C A A C A C T G C C C T C C C C A A A A G A A G C C C A A A C T C A A G G A G A A G T G C G 367

OTC

```

410  GlnAlaIglYanCylsYsYrProLeuGlnIlyAspIlyLeuProIlyuStLe 426
426  uueSerAspIleuSerLeuGlnIserLeuThraIleuThraIleuGlnIValG 443
443  IuAnIleSerAsnThrValGlnGlnIleuLeuLeuSerLysAlaIValA 459
459  GGCCTCCGACCTCAGCCTCTCAGAGACCTCCAGAGCGCTACCTCCAGAG 105
106  AAAACATCTCCAAACACCGGACGAGCGCTTCCTGCTCCAAAGCTACCAAG 155
146  ProGlnIlyLysIglYalYalYasIleuValSerAArgThrProGlnGlnI 476
156  CCGCAAGAAAGAAAGGGGACGACAGACCTGCTCCAGGACCTCCAGAGAGCA 205
476  sIysSerGlnIHisCysSerProGlnIglYserGlyThrSerAlaIleuProA 493
206  CAAAGCCAGCACTGTAGCCCTGAAGGAGACGGCTACTCAGCTACGACCGAG 255
493  IaGlyThrProIleuSerGluProProSerSerThrProGlnIserThrHis 509
256  CGGGACACACCGCTGTCTGACGCGCGAGCAGCAGCGCAAGCTCACACCCAT 305
510  AlaGluProGlnIleuAlaAspIlyrLeuSerGlySerGluAspProLeuG 526
306  GCTTAGCGCAGAGAGACGCTGACTTACTGATGAGTGGCTCTGAGAGACCGG 355
526  uArgSerPheIleuTrpCysAlaGlnAlaAargIylSerProAlaArgVal 542
356  GCGGACACTGGCTCTACGACGACGAGCGCGGAGACCCCGCGCAGGTATAC 405
543  AsnSerAsnSerLysAlaIlyProGlnIleuSerLeuThrCysSerValTrh 559
406  AACAGCAACATCCCAAGCGTAAGCCCGAGCTGTGTGTCACCTGCTCCTGTG 455
559  rSerProAspAspMetSerThrLysSerAspAspSerPheGlnIserLeuH 576
456  CTTCACCTGAGAGAAAGTTCACAAAGTTCGAGAGAGTCTCTCCAGAGCGTT 505
576  IaGlySerLeuProIleuAspSerPheSerLysPheValAlaGlyIuArg 552
506  CAGTAACTGCGCCCTTGAGAAGGCTCTCCGAAATTTGTGGGA.....AGG 549
593  AspCysProArgIleuLeuLeuSerAlaIleuIaGlnIlyAspIleuAlaSe 609
550  AACGGGAATTCGCGGTGTGGTCTCAATTCCTTGAGACAGAAAGATATGCCCG 599
609  rGluIleuGlnIlyLeu 614
600  GGAAT...CTGGAGATTG 612

```

QY	5755	taacagccacagctcgacccggacaaagagagccacatcgccacacagctgccggcgacctc	5814
Db	14859	taaacgcgcacagctcgacccggacaaagagagccacatcgccacacagctgccggcgaccttc	14918
QY	5815	ccgagagacgcgcagagacgcctactgcctgcgtgcagtcgacggagagatctgggagcgagacgc	5874
Db	14919	ccgagagacgcgcagagacgcctactgcctgcgtgcagtcgacggagagatctgggagagacgc	14978
QY	5975	ccacagccgaaacagggctgcacaaacatgagtcgagagagagggggtccggccacagcccgcgcg	5934
Db	14779	cccaaccgaaacagggctgcacaaacatgagtcgagagagagggggtccggccacagcccgcgcg	15038
QY	5935	ggagagccacagagagacacatcgtgtgcatacgacgcctgcgcgtgcctgcagtcgagaccgagcgcctactac	5994
Db	15039	ggagagccacagagacacacatcgtgtgcatacgacgcctgcgcgtgcctgcagtcgagaccgagcgcctactac	15099
QY	5995	ggtagccggagagacatcctttggctgcgcag	6022
Db	15098	ggtagccggagagacatcctttggctgcgcag	15125

RESULT	4
LOCUS	AB058723 5915 bp mRNA PR1 05-JUN-2001
DEFINITION	Homo sapiens mRNA for KIAA1820 protein, partial cds.
ACCESSION	AB058723
VERSION	AB058723.1 GI:14017856
KEYWORDS	
SOURCE	Homo sapiens brain cDNA to mRNA, clone:h01321.
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (sites)
REFERENCE	Nagase,T., Nakayama,M., Nakajima,D., Kikuno,R. and Ohara,O. Prediction of the coding sequences of unidentified human genes. XX The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro DNA Res. 8 (2), 85-95 (2001) 2 (bases 1 to 5915) Ohara,O., Nagase,T. and Kikuno,R. Direct Submission Submitted (27-MAR-2001) Osamu Ohara, Kazusa DNA Research Institute, Department of Human Gene Research; 1592-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:cdna@f.kazusa.or.jp). URL:http://www.kazusa.or.jp/huge, Tel.81-438-52-3913, Fax:81-438-52-3914)
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	
REFERENCE	
FEATURES	
source	Location/Qualifiers 1..5915 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="h01321" /tissue_type="brain" /note="vector:pBluescriptII SK plus" 458..5392 /gene="KIAA1820" 458..5392 /name="KIAA1820" /note="Start codon is not identified." /codon_start=1 /product_id="BAB47449.1" /db_xref="GI:14017857" translation="PARVOSPERERGPHGKQNVGYOQTSGTSLNRYNQPSQAQLSGD RGLADSSPIPRFTHAGESSLDANGAGQPPEPPQPLPAGVAKTIDENMTKTAVPSRS IAGGAVERPRHSILHYVOPPPQPLAPYPILOKRIQMDIASLFPQGTFFPOHQ SEPTSTSYSSVVGOGGASYSKCTAAPBHRPLATSSSIAGAROYNLHAVVGGS RLSTDQQGQGGQGGQGGALORSHAOETLHTONAKIKOHGGOOGGCGDDAAVTG PEQYYTFSSSSSPARSYGSRPSYSTPEMLPLNFEPYSGOOITSGCARPAETID HSHMPFNPSPTLASVDYTOGNKKELQMDIPHLNISLSLSQSITLVLTJVSVENTS NTVOQLLKSAIVPKRGKVKNLVSRFPQRHSDHCSPREGSITSAPRPGLISEPSSST
CDS	
gene	

POSTHAEPEADYIAGSEDEPLERSTVTCANARSPARVNSMKAKESEVTCVTSAD  
DMSTKDDSDRSLJGSDPLDPSKSVAGNEDCPILLSLALOGJALBETAL  
KAKRAVAEPLSVIOWSKRPPRELEHNSHSCIDYAKSAMPRGEBALPSPLODTTAN  
AKPSGTELEDBESVAFTDPDKATTTGPTGPTLGTATPILGTAPADALFPCEDPHTTAS  
SASASNAPFMEBENTGDACPFRWGLPGLBETLGLMOGAGADGJISKDTHESALCGRQ  
EEDPDEKTAJLPDFFKOEYGVGVEENAGJLQCCPVAKADRWLEDSRHCSTADPBD  
LPIPTSKREDELAEREYSJLCELLGSPDEPRGMDPLSJKAPLICTEHEVENDLS  
KAAGSGPCLSGESVITLLPGTYETBSKYQMSFNSJLHKKGEEDGPGKARAPAGSTIS  
DASLQAKPRAVPEBALAKKPRVPRGKSLRSRRVRGLPBAEBSPPALVPLKOLL  
PESTQEPQOGMEAGAPGRGASBEUPMKTREBTLTASBPPTPEPGLVTPAPDPLD  
GGKORAFATSOGRVKGKSPFLAASSPSNAPLAPYASDSBPMKSSKTMEDSPSPGKDOOR  
SMILSRFTYOLFEHSKRKRSPBGBRLPWCNRATKLINDSHLPAFTYRSSPOKAGRSY  
ORAVPEPDSGLKSLSDPLTALKRKRSAPMAVYPTKRLNIVARSSSSSSNAGNAGGCG  
KEEBSPEPDLKESRPAKTPKNGEPEFTKILPEPPEPACAKILASRAQGMNT  
KYLDPKRGSGIKLAIATYOKTITPSLCLPACKACAPASAPGNDLPSLSDPOKTEATSSY  
FVGLGEBALVAVTGTAKLPTSGADPLCLRNPTPTRLKGLLNSKSLTSPDPSGAGASSET  
BALPQGBTALAPKRSRGRGAGAHGISGPELEKXPVPLGAPLILPDRASGNOGAKSEB  
NSGGGKGRKEMELGIALSQBPSPGQVQYQAKQOPEHTNYSSSKNRPITGGRKLRN  
TSSCGKRAKRRQOQVLPDPAPEPIRTKLTISCKRSLSDSKPLTSPLESLDQAK  
PFSGONAPPADADARPLCTTDRRAVSAEBOG"

[illegible][illegible]

[illegible]

RESULT	5				
AX024201					
LOCUS	AX024201	5395 bp	DNA		
DEFINITION	Sequence	1 from Patent WO008143.		PAT	15-SEP-2000
ACCESSION	AX024201				
VERSION	AX024201.1	GI:10184512			
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi				
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 5395)				

FEATURES		source	
CDs	1. >5394 /note="unnamed protein product" /codon_start=1 /protein_id="CAC08943.1" /db_xref="GI:10184513"	1. 5395 /organism="Homo sapiens" /db_xref="taxon:9606"	1. >5394 /note="unnamed protein product" /codon_start=1 /protein_id="CAC08943.1" /db_xref="GI:10184513"
Location/Qualifiers	1. 5395 /organism="Homo sapiens" /db_xref="taxon:9606"	1. 5395 /organism="Homo sapiens" /db_xref="taxon:9606"	1. 5395 /organism="Homo sapiens" /db_xref="taxon:9606"
Authors	Poustika, A., Lehrich, H., Radolof, J., and Seranski, P.	Poustika, A., Lehrich, H., Radolof, J., and Seranski, P.	Poustika, A., Lehrich, H., Radolof, J., and Seranski, P.
Title	Gene isolated on the short arm of human chromosome 17	Gene isolated on the short arm of human chromosome 17	Gene isolated on the short arm of human chromosome 17
Journal	Patent: WO 0008143-A 1 17-FEB-2000; DEUTSCHE KLOBESFORSCH (DE) ; POUTSKA ANNEMARIE (DE) ; RAELOF UWE (DE) ; SERANSKI PETER (DE) ; MAX PLANCK INST FBER MOLEKULAR (DE)	Patent: WO 0008143-A 1 17-FEB-2000; DEUTSCHE KLOBESFORSCH (DE) ; POUTSKA ANNEMARIE (DE) ; RAELOF UWE (DE) ; SERANSKI PETER (DE) ; MAX PLANCK INST FBER MOLEKULAR (DE)	Patent: WO 0008143-A 1 17-FEB-2000; DEUTSCHE KLOBESFORSCH (DE) ; POUTSKA ANNEMARIE (DE) ; RAELOF UWE (DE) ; SERANSKI PETER (DE) ; MAX PLANCK INST FBER MOLEKULAR (DE)
BASE COUNT	1117 a 1877 c 1595 g 806 t	1117 a 1877 c 1595 g 806 t	1117 a 1877 c 1595 g 806 t
ORIGIN	76.8%; Score 4627.6; DB 6; Length 5395; Best Local Similarity 92.7%; Pred. No. 0; Matches 5131; Conservative 0; Mismatches 9; Indels 398; Gaps 10;	76.8%; Score 4627.6; DB 6; Length 5395; Best Local Similarity 92.7%; Pred. No. 0; Matches 5131; Conservative 0; Mismatches 9; Indels 398; Gaps 10;	76.8%; Score 4627.6; DB 6; Length 5395; Best Local Similarity 92.7%; Pred. No. 0; Matches 5131; Conservative 0; Mismatches 9; Indels 398; Gaps 10;
Query	490 atgcagctcttcgcagaaagctctggttcacatggaacacagaaactaccagagacc 549	490 atgcagctcttcgcagaaagctctggttcacatggaacacagaaactaccagagacc 549	490 atgcagctcttcgcagaaagctctggttcacatggaacacagaaactaccagagacc 549
Db	1 ATGCAAGCTCTTTCGAGAAAGGTGTGGTTCATAGCAAAACAAGAACTACCAAGACAGCC 60	1 ATGCAAGCTCTTTCGAGAAAGGTGTGGTTCATAGCAAAACAAGAACTACCAAGACAGCC 60	1 ATGCAAGCTCTTTCGAGAAAGGTGTGGTTCATAGCAAAACAAGAACTACCAAGACAGCC 60
QY	550 tcgcaggaacaacacacgcctcagaagattaccagagacccagatcagagccgagctacgtgc 609	550 tcgcaggaacaacacacgcctcagaagattaccagagacccagatcagagccgagctacgtgc 609	550 tcgcaggaacaacacacgcctcagaagattaccagagacccagatcagagccgagctacgtgc 609
Db	61 TCGAGGAACACACAGGCTTAGAAGATTACAGGAGCCAGAGTCAAGCGGGGCTTAAGCTGC 120	61 TCGAGGAACACACAGGCTTAGAAGATTACAGGAGCCAGAGTCAAGCGGGGCTTAAGCTGC 120	61 TCGAGGAACACACAGGCTTAGAAGATTACAGGAGCCAGAGTCAAGCGGGGCTTAAGCTGC 120
QY	610 gaccgcagcagcgtctctgcacagagacattataaccgcagacctatacccgagctatgag 669	610 gaccgcagcagcgtctctgcacagagacattataaccgcagacctatacccgagctatgag 669	610 gaccgcagcagcgtctctgcacagagacattataaccgcagacctatacccgagctatgag 669
Db	121 GACCGGACGGGCTCTGCGCAAGGACATTAATAACCGCAGCCTTACCGCGAGCTAATGAG 180	121 GACCGGACGGGCTCTGCGCAAGGACATTAATAACCGCAGCCTTACCGCGAGCTAATGAG 180	121 GACCGGACGGGCTCTGCGCAAGGACATTAATAACCGCAGCCTTACCGCGAGCTAATGAG 180
QY	670 gttggcctcggagcgcctctgcagctcagcgcggttggccgcgcagaagtataccacgga 729	670 gttggcctcggagcgcctctgcagctcagcgcggttggccgcgcagaagtataccacgga 729	670 gttggcctcggagcgcctctgcagctcagcgcggttggccgcgcagaagtataccacgga 729
Db	181 GGTGAGCTCGGACAGCCCTCTGCACTACAGCCCGGCTGGCCGCGACAAAGTACACCGCA 240	181 GGTGAGCTCGGACAGCCCTCTGCACTACAGCCCGGCTGGCCGCGACAAAGTACACCGCA 240	181 GGTGAGCTCGGACAGCCCTCTGCACTACAGCCCGGCTGGCCGCGACAAAGTACACCGCA 240
QY	730 ggcagcagagccctccacacagagagcctcgaagggagagccgaggttcctccgtggtac 789	730 ggcagcagagccctccacacagagagcctcgaagggagagccgaggttcctccgtggtac 789	730 ggcagcagagccctccacacagagagcctcgaagggagagccgaggttcctccgtggtac 789

Resid. No.	Score	Match	Query Length	DB	ID	Description
1	6022	100.0	6022	6	AX004680	AX004680 Sequence
2	5561	99.0	160544	2	HS253P07	AL535490 Homo sapi
3	5901.2	98.0	22843	9	HS2711791	AL771191 Homo sapi
4	4752.2	79.1	5915	9	AB058723	AB058723 Homo sapi
5	4627.6	76.8	5395	6	AX024201	AX024201 Sequence
6	4627.6	76.8	5395	6	AX024201	AX024201 Sequence
7	3664	60.8	7222	10	MTSMA	AL771790 Homo sapi
8	970.4	16.1	2660	9	HS8801433	D29801 Mouse mRNA
9	495	8.2	595	11	AF021115	AL133649 Homo sapi
10	391.8	6.5	495	11	AF021118	AF021115 Homo sapi
11	326	5.4	542	11	G09710	AF021118 Homo sapi
12	317	5.3	330	6	AX004677	G09710 Human SRS C
13	206.4	3.4	328	6	AX004676	AX004677 Sequence
14	110	1.8	6542	9	AB006630	AX004676 Sequence
15	110	1.8	43761	9	HS59752	AB006630 Homo sapi
16	101.2	1.7	6212	10	AF007594	AL031346 Human DNA
17	85.4	1.4	1512	10	MMU70655	AY007594 Mus muscu
18	85.4	1.4	1512	10	MMU70655	U70655 Mus muscu
19	85.4	1.4	14625	9	NC88RYLC	U70656 Mus muscu
20	83.6	1.4	303091	2	AC084799	X67204 Mus muscu
21	82.2	1.4	10266	22	E11536	AC084799 Mus muscu
22	82	1.4	87076	9	AC005918	E11536 DNA from y
23	81	1.3	298166	2	AC087563	AC005918 Homo sapi
24	79	1.3	152547	2	AL137857	AC087563 Homo sapi
25	78.8	1.3	197332	2	AC073946	AL137857 Homo sapi
26	78.6	1.3	1611	10	MMU70651	AC073946 Mus muscu
27	78.4	1.3	200542	2	CNS01R83	U70651 Mus muscu
28	75.8	1.3	167390	9	AC007263	AL157736 Homo sapi
29	75.4	1.3	1549	10	MMU70657	AC007263 Homo sapi
30	75.4	1.3	178343	2	AC024049	U70657 Mus muscu
31	74.8	1.2	1968	10	MMU46463	AC024049 Homo sapi
32	74.4	1.2	162268	9	AC000097	U46463 Mus muscu
33	74.4	1.2	168239	9	AC007663	AC000097 Homo sapi
34	74.4	1.2	174840	9	AC006549	AC007663 Homo sapi
35	74.4	1.2	179269	9	AC006547	AC006549 Homo sapi
36	74.2	1.2	41185	9	HSB11725	AC006547 Homo sapi
37	74.2	1.2	18119	9	AL161726	AL043845 Homo sapi
38	73.8	1.2	238652	2	AC044845	AL161726 Human DNA
39	73.6	1.2	1547	10	MMU70649	AC044845 Mus muscu
40	73.4	1.2	1546	10	MMU70642	U70649 Mus muscu
41	73.4	1.2	1546	10	MMU70645	U70642 Mus muscu
42	73.4	1.2	1550	10	MMU70641	U70645 Mus muscu
43	73.4	1.2	1552	10	MMU70650	U70641 Mus muscu
44	73.4	1.2	2298	10	AF337043	U70650 Mus muscu
45	73.4	1.2	2319	10	AF337045	AF337043 Mus muscu
46	73.4	1.2	2319	10	AF337045	AF337045 Mus muscu



ORIGIN									
Query Match 100.0%; Score 6022; DB 6; Length 6022; Best Local Similarity 100.0%; Pred. No. 0; Matches 6022; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	ggatccagcagcccaagggtgagggagcgaaattctctgtctataa-gcttttgagc	60						
Db	1	GGATCCAGCAGCGCCCAAGGGGATGAGGAGCGGAAATTCCTGCTAAATGCTTTTGAGC	60						
QY	61	tgfcagaaagggctggagtgatggtgggagacattgggagagctggcaatggcgg	120						
Db	61	TGFCAGAAAGGGCTGGAGTGATGGGTGGGGACATTGGGAGAGCTGGCAATGGCGGG	120						
QY	121	gggggggagggtagctcccaagtagctggcgtggcgagcggttttggctcccgat	180						
Db	121	GGGGGGGGGAGGTAGCTCCCAAGTAGCTGGCCTGGGAGCGCGGTTTTGCTCCCGCAT	180						
QY	181	cagtggccgtctctggcaagctcagctcgaggcgatgtggagcgggaaattacagagca	240						
Db	181	CAGTGGCCGCTCTCTGGCAAGCTCAGCTGCAGCGATGTGGAGCGGGAAATTACAGAGCA	240						
QY	241	caactccctgacacagaagttgcaatatgcacagctggtgggaggtccagcgaa	300						
Db	241	CACCTCCCTGACACAGAAGTTGCAATATGCGCACAGCTGGTGGGAGGCTTCAGGCGAAG	300						
QY	301	gggggactattagagctgcgcggggagcagcgaggtgggaggtgggaggggtg	360						
Db	301	GGGGGACTATTAGAGCTGCGCGGGGAGCAGCGGTGGGAGGTGGGTGGGAGGGTG	360						
QY	361	ctttctgagcaaaagaaagtggccggtctgaatcgctcatctctgcccctccctg	420						
Db	361	CTTTCTGAGGCAAAAGAAAGTGGCCGCTGAATGCGTCATCTCTGCCGCCCTCCCTGGC	420						
QY	421	catctccctccctccctccctccctccctccctccctccctccctccctccct	480						
Db	421	CATCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT	480						
QY	481	gcccagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag	540						
Db	481	GCCCGAGTCATCAGCTCTTTTGAGAAAGGTGTGGTTTCATGTGCAAAACAAACAACTAC	540						
QY	541	cagcagacctcgagcaaaacatcacctcagagattacagcagcagcagcagcag	600						
Db	541	CAGCAGACCTCGCAGAAACATCACGCTTAGAGAAATTACAGCGCGGAGTCAGCGCGGG	600						
QY	601	ctaagctggagcggcagcgctgctgcgcaagagctattataaccgagccttaaccg	660						
Db	601	CTAAGCTGGCAGCGGACGCGCTGCTGCGCCAGAGNCTATTATACCGCGGCTTACCCG	660						
QY	661	agctatgaggtggcgtggcagcgcctctggcactgagcgcggtggcgcgcgcaag	720						
Db	661	AGCTATGAGGTTGGCGCTGGCAGCGCCTCTGGCACTGACGCGGCTGGCGCGGCAAG	720						
QY	721	tacacocgaggaagcaagccctgcccacacagcaagccctgaggggagcggttttc	780						
Db	721	TACCACCGAGGACAGAGGCGGCTGCCCCACACAGAAAGGCTGCAGGGAGGCGGCTTC	780						
QY	781	cttgctcagcgctcaggaacagcagcagcagcagcagcagcagcagcagcagcag	840						
Db	781	CTTGCTCAGCGCTCAGGACGAGCGGCTTACCGAGCGGCTATGCTGGTGGAGAGAGC	840						
QY	841	cttcaggtctggggggcccccagcagcagcagcagcagcagcagcagcagcagcag	900						
Db	841	CTTCAGGCTTGGGGGGCCCCACAGCAACACACACACACACACACACACACACACAC	900						
QY	901	gtggcgaagtatgatgagaacttgatgaaaaagacagcagtgccccccagcagcag	960						
Db	901	GTGGCCGAAGTATGATGAGAAGTTGATGAATAAGACAGCAGTGGCCCCCAGCAGCAT	960						
QY	961	gpagagcagggcccgaggtgcccctctcgagcactccctcgacgtccagcagccacg	1020						
Db	961	GTGAGCGGCTCCGAGGAGCCACTGTGAGCGCAGCTTCTCTACTGCAACAGGCGCGTGGC	1020						

Db	961	GCAGAGCAGGGGCGCCAGGTGCCCTTTGGGACTACTCCTCTGCACGCTCAGCAGCCACG	1020
QY	1021	cggccccagcagcccccctggcatacccaagctcccaaggcagaagctgcagaacgacatt	1080
Db	1021	CGGCCCCAGCAGCCCTGGCATACCCCAAGCTCCAAAGGAGAGCTGCAGAGCAGCATT	1080
QY	1081	gctcccccctgcccctcccccagggtaccacttctcagcattccagtccttcccc	1140
Db	1081	GCTTCCCTCTGCCCCCTTCCCCAGGTACCCACTTTCCTCAGCATTCCTCAGTCTTCCC	1140
QY	1141	ecctctccactactcctctctcaggggtggtgggagggggcccaactcctataag	1200
Db	1141	ACCTCTCCACTACTCTCTCTGTCCAGGGTGTGGGAGGGGCCCCACTCTCTATAAG	1200
QY	1201	agttgacagcagccagctgcccagccccatgacagggcgctgactgccaagctccagctg	1260
Db	1201	AGTTGACAGCAGCCAGCTGCCAGCCCCATGACAGGCGCTGACTGCCAGCTCCAGCGCTG	1260
QY	1261	gccccggggcagcgggtccagaatcttcactaccagtcggggcgccctcagctatgac	1320
Db	1261	GCCCCGGGGCAGCGGTCAGAAATTTCTATGCTTACCAGTGGGGCGCTCAGTATATGAC	1320
QY	1321	cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag	1380
Db	1321	CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	1380
QY	1381	catgccaggaacccctccattaccaaaaacctcgccaaagtatcagcaactaaggggcagaa	1440
Db	1381	CATGCCAGGAACCCCTCCATTACAAAAACCTCGCCAAAGTATCAGCACTACGGGAGCAA	1440
QY	1441	ggccagggctactgcccagccgagcagcagcagcagcagcagcagcagcagcagcag	1500
Db	1441	GGCCAGGGCTACTGCCAGCCGGAGCGCGCTCCGGAGCCCGGAGAGGAGTACACAGACC	1500
QY	1501	ttcagccccagctccagcagcagcagcagcagcagcagcagcagcagcagcagcagcag	1560
Db	1501	TTCAGCCCCAGCTCCAGCACTCACCGCGCGCTCGGTGGCGGCTCACCTTCTCTACAGT	1560
QY	1561	tcacacagctgcgcgtgatgcacacacacacacacacacacacacacacacacacacac	1620
Db	1561	TCCACAGCTGCGCGGTGATGCCAAACCTGGAGAACTTTCCCTACAGCCAGCAGCGGTC	1620
QY	1621	agcagcggccctcccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcag	1680
Db	1621	AGCAGCGGGCCCTTCCCGCAGGGATCAGTACACAGCCAGCCACTTCAAGCCCTGCTCAAT	1680
QY	1681	ccctcccaacagctgac	1740
Db	1681	CCCTCCCAACAGGATGCCACAGCTGTGGACACACAGGCTGGCAACTGCAAGCCCTT	1740
QY	1741	cagaagcaagctccctgagac	1800
Db	1741	CAGAAGCAAGCTCCCTGAGAACTCTGTGCGATCTCAGCCTGCAGAGGCTCACGGCG	1800
QY	1801	ctgaccttcagctgagaaac	1860
Db	1801	CTGACCTTACAGTGGAGAACATTCACACACCGCTCCAGCAGCTGCTGCTCTCCAGGCT	1860
QY	1861	gctgtgcgcagagaagagtgatcagaacacacacacacacacacacacacacacacacac	1920
Db	1861	GCTGTGCGCAGAGAAAGGTGTCAAGAACTCTGTCCAGGACCCCGAGAGCAGCATAAA	1920
QY	1921	agcagcaactgcagccccgaaggagcggtactcagccagcagcagcagcagcagcagcagc	1980
Db	1921	AGCCAGCACTGCAGCCCCGAAAGGAGCGGTACTCAGCCGAGCGCCCGAGGACACCCGCTG	1980
QY	1981	tcagagcccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag	2040
Db	1981	TCAGAGCCCGCGAGCAGCAGCGCACGCTCCACGATCCAGGAGCCCGAGGAGCCGAGCTAC	2040
QY	2041	ctgagcggtcccgaggaacccactgagcgagcttctctactgcaaccagggccggtggc	2100
Db	2041	CTGAGCGGCTCCGAGGAGCCACTGTGAGCGCAGCTTCTCTACTGCAACAGGCGCGTGGC	2100



QY	2101	agcccttcgagggtaataaagaactctgaagggcagaagccgaatcagctgtccactcttct	2160
Db	2101	agcccttcgagggtaataaagaactctgaagggcagaagccgaatcagctgtccactcttct	2160
QY	2161	gtgacacctctctgaagacatctccacccaattctagacatccctccagaagctctacacgyc	2220
Db	2161	gtgacacctctctgaagacatctccacccaattctagacatccctccagaagctctacacgyc	2220
QY	2281	ctgctcaagcgccctctgacagagagacactgagctccgagatccctctgaggtctbcaagagac	2340
Db	2281	ctgctcaagcgccctctgacagagagacactgagctccgagatccctctgaggtctbcaagagac	2340
QY	2341	atcgtgtgaagaagcgacaaagcttgggtctgaagacaccaacgcctggtcagaagagacagc	2400
Db	2341	atcgtgtgaagaagcgacaaagcttgggtctgaagacaccaacgcctggtcagaagagacagc	2400
QY	2401	aagcgacacctctcgtctgtgaataacacaagcgctgacctgtgaactctgtgccaagaagtgs	2460
Db	2401	aagcgacacctctcgtctgtgaataacacaagcgctgacctgtgaactctgtgccaagaagtgs	2460
QY	2461	tggccccgagctctgggagacgagagacgctgcgcgacactcctctgacgctggaacaagggcgyc	2520
Db	2461	tggccccgagctctgggagacgagagacgctgcgcgacactcctctgacgctggaacaagggcgyc	2520
QY	2521	aatgcacaaagactcaagcccagggctgttttgaagaacctccgttgctctctgtaagcct	2580
Db	2521	aatgcacaaagactcaagcccagggctgttttgaagaacctccgttgctctctgtaagcct	2580
QY	2581	gaccccaaaaagacaactcgtgctccctccctctctgtgtacaaagcccaacctctgggtctct	2640
Db	2581	gaccccaaaaagacaactcgtgctccctccctctctgtgtacaaagcccaacctctgggtctct	2640
QY	2641	gctccagaacctccaactacagacagcttlttgacgtttcccgagacaacacgctgacagatca	2700
Db	2641	gctccagaacctccaactacagacagcttlttgacgtttcccgagacaacacgctgacagatca	2700
QY	2701	ggagagacagcgcaaacaccttgcctctgagcagaagagaaacctctggagatgagcttgcacag	2760
Db	2701	ggagagacagcgcaaacaccttgcctctgagcagaagagaaacctctggagatgagcttgcacag	2760
QY	2761	tggggaattgcacacctctgcgagacttaaccaagggcctctgagacaggtgtggaaagacccaaga	2820
Db	2761	tggggaattgcacacctctgcgagacttaaccaagggcctctgagacaggtgtggaaagacccaaga	2820
QY	2821	ggatcaccagaaggggacacccaatgagcttcgagctgcgtgcgtgcgtccagaagagagagac	2880
Db	2821	ggatcaccagaaggggacacccaatgagcttcgagctgcgtgcgtgcgtccagaagagagagac	2880
QY	2881	cccccctggagagaaagttgcgctctgtcccccggagacattcaagaacagagagaggtgtgtgtg	2940
Db	2881	cccccctggagagaaagttgcgctctgtcccccggagacattcaagaacagagagagaggtgtgtgtg	2940
QY	2941	gt'gaaagagagagcaggtgtggctgtgcagtgagcccgagaggtgtgcgaaggtgcacggtgtg	3000
Db	2941	gt'gaaagagagagcaggtgtggctgtgcagtgagcccgagaggtgtgcgaaggtgcacggtgtg	3000
QY	3001	ctggagagacagcgagactcgtcttcacacgcgcagcttgggggaacttcgaactcctctgcgca	3060
Db	3001	ctggagagacagcgagactcgtcttcacacgcgcagcttgggggaacttcgaactcctctgcgca	3060
QY	3061	cccaacacagagagagagacacttgaaagcttgagaggtgagatcctccctatctgtgagctc	3120
Db	3061	cccaacacagagagagagacacttgaaagcttgagaggtgagatcctccctatctgtgagctc	3120
QY	3121	ctggagagacaccccgagagagagcctggtatgtacagagcccgctgtgcacccaagggcccaactc	3180
Db	3121	ctggagagacaccccgagagagagcctggtatgtacagagcccgctgtgcacccaagggcccaactc	3180

QY	3181	atctgcacacaaagagagctgagagagctgcctgactccaagccgccttgagcctctccg	3240
Db	3181	ATCTGCACCAAGAGAGAGTGTGAGAGAGTGTGGATCTCCAAAGCCGGCTGTGGCTCTCCG	3240
QY	3241	tgcacacctccaaagagatccgtatccctgcctgcgcctcaagctgagccagagccaag	3300
Db	3241	TGCACACTTCCAAAGGAGATCCGTATCCCTGCTGGGCCCTTACGTGGCACCCAGTCCAAAG	3300
QY	3301	gtccagagctggtttgagatccctctcgttccacaatgaagccagctgaagaggggcctgat	3360
Db	3301	GTCCAGAGAGCTGTTTGAAGTCCCTGCTGTCCACATGAAAGCCAGTGAAGAGGGGCGTGAAT	3360
QY	3361	ggggagagagctccaaagagattccacacctccgaagcgtctctctgcgccagaagccaac	3420
Db	3361	GGGAGAGAGCTCCAAAGGAGATTCCACACCTCTGAGACCTCTCTGTGGCCAGAGGCCAAC	3420
QY	3421	aagagctgctgtgcacaggggccccatccgaagaagaagagctgtgcacaggggccaagagc	3480
Db	3421	AAGCTGCTGTGCTCCAGAGGCCCATCCGAAAGAAAGAGCTGTGTCCACAGGGGCCAAGAGC	3480
QY	3481	ttacagagacgctgcgtgtgcacacgggggctgcgcagagccgagagactcccaatgcagggca	3540
Db	3481	TTACAGAGACGCTGCGGTGTGACACGGGGGCTGCCAGGCCAGAGACTCCCATGTCAGAGCA	3540
QY	3541	ccagctgctgcgccaaagacctttcttccctgaatctgcacagaggtcccccagggagag	3600
Db	3541	CCAGCTGCTGCCAAAGACCTTTCTCTCTGAACTGTGCACAGAGGCCCCCCCAGGGAGAG	3600
QY	3601	atcgaaagaggtgtagaaccccaagggcgcggggagctccggaagaaggtccccaagatgtact	3660
Db	3601	ATCGAAGAGGTGTGAAGCCCAAGGCCGAGGGGGGCTCCGAAAGAGGTCTCCAGAGATGTACT	3660
QY	3661	cgctatctcaagggccctgtgtagagcccccgcagccgcgagcccgagaccgcagctgacacacac	3720
Db	3661	CGCTATCTCAAGGCCCTGTGTAGAGCCCCGCGAGCCCGGAGCCCGGAGCCGCTGACACACAC	3720
QY	3721	ccctgacccccagacaacactggggggcagaagcagcgcccttaagtctggcgaagcgg	3780
Db	3721	CCCTGACCCCCAGACAACTGGGGGGGTAAGCAGCAGCCGCTTCAAGTCTGGGCAAGCGG	3780
QY	3781	gtggggaagccctccacaaagagctgcctccagcccccgaacccgggcgcgcctgtgactg	3840
Db	3781	GTGGGGAAGCCCTCCACAAAGAGCTGCTCCAGCCCCGACAAACCCGGCGCCTCTGTGAG	3840
QY	3841	gctcccgacagccagcccgatgtgtctccaaagacaaaggagacagactccacccagacgcct	3900
Db	3841	GCTCCGACAGCCAGCCCGATGTGTCTCCAAAGACAAAGAGAGACACTCACCCAGACGCCT	3900
QY	3901	ggcaagagccagcgtctcatatcatcttcggtcagcgcacaaaccacccagagagatctccac	3960
Db	3901	GGCAAGAGCCAGCGTCTCATATCATCTTCCGTTACGCACCAAAACCCAGAGACTTCCAC	3960
QY	3961	tccaagcggcgagagccctctgagagggcggtcccccacatcgccgtgtgcacaagaagcttc	4020
Db	3961	TCCAAGCGGAGAGCCCTCTGAAGGGCGGCTCCCAACTCGCTGTGCACCAAGAAGCTTC	4020
QY	4021	ctcgaacacagccacactgtgcgcgacaattcaagagctctccagcagcccccaagaagagggc	4080
Db	4021	CTCGAACACACCCACTTGTGCCCGCACATTCAAGAGCTCTCAGCAGCCCCCAAGAAGAGGC	4080
QY	4081	aaggctgagccagcgggcagaggttccccaacactgtgtcgaagcagcagactctctgcacgg	4140
Db	4081	AAGGCTGAGCCAGCGGGCAGAGGGTCCCAAACCTGTGTGAGGAGCAGCAAGCTCTTGACCGG	4140
QY	4141	ccccctcatggtctcaaaagagatctggtcttcacatgggcggttccccaagaagaagcgg	4200
Db	4141	CCCCCTCATGGTCTCAAAAGAGATGTCGCTTCATGAGCGCGGTCCCCACCAAGAAGCGG	4200
QY	4201	aaactgtgcttggcgacggcagagcagagacagcagcagcagctgccaattggggagagt	4260
Db	4201	AAACTGTGCTTGGCGACGGCAGAGCAGAGCAGACCAACCTGCAATGGGGAGAGAGT	4260
QY	4261	ggaagagagagagagcctgtaggtctccacaccccttcaagagagatgtctcttccacaaga	4320



1

QY 1621 agcaccggggtccctccgagggatcaactgaccacagcaacttcacgtccctgctcaat 1680  
Dd 108719 AGCACCGGGGCTTCCCGCAGGAGTCACTGACACACACCACTTCACTGCCCCCTGCTCAAT 108660  
QY 1681 cctccccaacagatgaccacagctctgtggacacccagcgtggaactggaagccctt 1740  
Dd 108659 CCTCCCAACAGATGCCACACAGCTCTGTGGACCCACAGGTGGCACTGCAAGCCCTT 108600  
QY 1741 cagaagacagctccctgagaactgctgtcggatctcagctgcagagctcagcgcg 1800  
Dd 108599 CAGAAGACAAAGTCCCTGAGAACCTGCTGTGGATCTCAGCTGCAGAGCCTCAAGCG 108540  
QY 1801 ctgaccttacagtggaagaaatctcaaacagctcaagcagctgtgtctcacaagct 1860  
Dd 108539 CTGACCTCAGAGTGAGAAATCTCCAAACCGCTCCAGCAGCTGCTCTCCAAAGCT 108480  
QY 1861 gctgtgcgcagagaagaggtgtcaagaactctgtgtccagacccacagcagcataaa 1920  
Dd 108479 GCTGTGCCGAGAGAAAGGTGTCAAGAACTCTGTGTCCAGGACCCCAAGACGATATA 108420  
QY 1921 agccagcaactgcagcccgagggagcgtactcagccgagccgcagggcacacgctg 1980  
Dd 108419 AGCCAGCACTGCAGCCCGAGAGGAGCGGCTACTCAGCCGAGCCGCGAGCACACCGTG 108360  
QY 1981 tcagagccgcagcagcaacgcacagctcaacagctgcagcagcagcagcagctac 2040  
Dd 108359 TCAGAGCCGCGGAGCAGCACGCCACAGTCCAGGATGTGGAGCGCGAGAGGCCGACTAC 108300  
QY 2041 ctgagcgctccagagacccactgagcgagcttctctactcaacagccgctgtgc 2100  
Dd 108299 CTGACGGCTCCGAGACCCACTGGAGCGCAGCTTCTCTACTGCACACCGCCGCTGC 108240  
QY 2101 agcctgcaggggtcaacagcaactcgaagccagccagctcgtgtccactgttct 2160  
Dd 108239 AGCCCTGCAGGGTCAACAGCAACTCGAAGCCGCAAGCCGCGAGTCCGCTGTCT 108180  
QY 2161 gtgaactctactacaaagtcaacaaatgtgacagactcttccagagctcaacgc 2220  
Dd 108179 GTGACCTCTCTGACAGATGTCCACCAATCTGAGAGCTCTCTCCAGAGCTTACCGGC 108120  
QY 2221 agctgcgcgtcagcagctctccaaagtctgtggcgtgagcgggactgtccggctg 2280  
Dd 108119 AGTCTGCCGCTCGACAGCTTCTCCAAATCTGTGGGGGTGAGCGGACTGTCCGGGCTG 108060  
QY 2281 ctgctcagcgcctggcagagaggaactgcctgcagatctcgtgggctgcaggaagcc 2340  
Dd 108059 CTGCTCAGCGCCCTGACAGAGGAGCTGCGCTCGAGATCTCTGGGGCTGCAGGAAGCC 108000  
QY 2341 atcgtgagaagccgcaaaaagcttggtgtgaagacccagcctgttcaagagacagc 2400  
Dd 107999 ATCGGTGAGAGGCGGACAAAGCTTGGGCTGAAGCACCCAGCCTGGTCAAGAGACAGC 107940  
QY 2401 aagccacccctctcgtggagaaacacagcgcctgctgactctgtggcagaagtgcg 2460  
Dd 107939 AAGCCACCTCTCTGCTGAGAACCCAGCGCTGCTGGACTGTGTGGCCAGAGTGC 107880  
QY 2461 tggcccccggctgggagccgagggccctgcagcactcctgcagctgagcagggcgcc 2520  
Dd 107879 TGGCCCGCGCTGGGAGCGGAGGCGCTGCGGACTCTCTGCACTGGGACAAAGGCGCG 107820  
QY 2521 aatgcagaagcttcagccagcgtgttgaagaccccttcgtggccttcactagcct 2580  
Dd 107819 AATGCCAAGGACTTACGCCAGGGCTGTGTGAAGACCCCTTCGCTGGGCTTCGCTAGCCT 107760  
QY 2581 gaocccaaaagaaactgtctctctctcttggtaacaaagccaccccttgggttct 2640  
Dd 107759 GACCCCAAAAGAACTGTCTCTCTCTTTGGTACCAAGCCACCCCTTGGGGTCTCT 107700  
QY 2641 gctcagacccactacagagctttgtactgtttccggagcaacacgcgtgcagctca 2700  
Dd 107699 GCTCCAGACCCCACTACAGCAGCTTTTACTGTGTTCGGGACCAACACCGCTGCAGCTCA 107640  
QY 2701 gcgagacgcgcaaccccttgcctggcgagagaaacacctgggggagctgtccaggg 2760

Dh 106559	GTGGGGAAGGCGCTCAACCAAGACGTGCCTCCAGCCACGAACCAACCGGCGCGCTTCGTG	106500
QY 3841	gctctccgacacgacgcccacatggtgctccaagaccagagagacacacccaagacgct	3900
Dh 106499	GCCTCCGACAGCAGCCCGCATGGGCTCCAAAGACCAAGAGACAACTCAACCCAGACGCGCT	106440
QY 3901	ggcaagacacgacgctccatgatctcttcgtcaagcaccaaaacccaagagatcttccac	3960
Dh 106439	GGCAAGGACACAGCGCTCCATGATCTTCCTGGTCACGACACAAACCAAGGATGTTCTCAC	106380
QY 3961	tccaaagcggcgagcgccctcagagcgcgctcccaactcgcgtctgcacaaagaatc	4020
Dh 106379	TCCAAAGCGCGAGGAGCCCTCTGAGAGCGGCGCTCCCAACTCCGTCGACCAAGAACTC	106320
QY 4021	ctcgcacacacgacacttgcccgccacattcaaggtctccagcgcgccacagaaagagggc	4080
Dh 106319	CTCGACAAACAGCCCATCTGGCCGCACTTCAGAGCTTCACAGACGCCCCAGAAAGAGGCG	106260
QY 4081	agagtgcgacgacgagcgaaggttcccacaaacttgtagagcagaaagtctctgcacgg	4140
Dh 106259	AAGGAGAACCAAGGGGCAAGGGGTCCCAAACTGGTAGGCAAGCAAGTCTCTGACCGG	106200
QY 4141	ccccccacatcgctccaagaagtgccgttcctacatgcgacggtccccaacaaagacgg	4200
Dh 106199	CCCTCTCATGCGCTCAAAAGAAATCGCTTCCTTAGGCGCGGCTCCCAACCAAGACGG	106140
QY 4201	aacctgctctgcgagcagcgacagcagcagcagcagcagcagcagcagcagcagcagc	4259
Dh 106139	AACCTGCTCTGGGAGCGGACGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	106080
QY 4260	gggaagagagagagcctcgaaggttcccaccaaccctctccaagagatgtctctcccaag	4319
Dh 106079	GGAAAGGAGGAGGCGCTGAGGTTCCCCACACCTTCCTCAAGAGGAATGTTCTCCACAG	106020
QY 4320	aaagcacaagcccccacaagagcgaatgcagagcctgcgacaaagctcccaaccccagagac	4379
Dh 106019	AAGGACAAAGCCACCAAGGCAATGCGAGACTGCGACAAAGCTCCACCCCGGAGACC	105960
QY 4380	ccc-attctctgcctcagaagctcgcctctc-ggcagccttcacagggggcattgaagccaa	4437
Dh 105959	CCCGATGCGCTGCCCAAGCTCGCTCTCGGCGAGCGCTTCCAGGGGGCAATGAAGACAA	105900
QY 4438	gtgctgcacaccccgaaagggccgggctcgaagctgcgaagccatcgtgcagaagatcaac	4497
Dh 105899	GTGCTGCGACCCCGGAGAGGCGCGGGCGCTTAACCTGGAAGCATCTGCGAGAAATCAC	105840
QY 4498	tgcgcacacccccaagaagttgcacatgcgaagcgccagggcgctctctctgtatctctg	4557
Dh 105839	TGCCCCAGCTCAAGAAATGTCGATGAAGGCCCAAGGGGCTCTCTGTATCTCTG	105780
QY 4558	agcccaactcccttcgcacaaagaacccgtgcgctcaaggtgcctgcgggcacgccaatggg	4617
Dh 105779	AGCCCATCCCTTTCCGACAAACCGTGGGCTCAAGGGTGTGTGGGGCAACCCAGTGGGG	105720
QY 4618	gtgaaagaagcctcgtlaaatgttgggcacgggcagaagctcccaactctcgggcgtat	4677
Dh 105719	GTGGAAGAGGCGCTGHAATATGTGGGCAACGGGCAAACTCCCAACTTCTGGGCGTAT	105660
QY 4678	cgcgtatgcgaaatccaacacaacgatcctttaaaggaaatccatgcgaacgtatgaagaa	4737
Dh 105659	CCGTTATCAAAATCCACCAACAGATCTTTAAAGGCAAACTCATGAACCACTAAGAAA	105600
QY 4738	ctctctctactgcctgtttcaaaacggagccttaactcccgagagggcctgcagcgcct	4797
Dh 105599	CTGTCTTCTACTGACTTTTCAAAACCGAGGCTTTACACTCCCGGAGGCGCTTCGAGGCT	105540
QY 4798	ggggagagactcctcgcgccttaagaagaagagcgcggaaagagcgcggcagggcccatgga	4857
Dh 105539	GGGGGAGACGCGCTGGCGCTTAAGAAAGAGAGCGGAGAAAGCGGACGAGGCGCCATGGA	105480
QY 4858	ctctccaaggcccgctgcgaagaaggcgccatcttggccgggctctgcctcgcctcgc	4917
Dh 105479	CTCTCCAAAGGCGCGCTGGAAGAGGCGCCATCTCTTGGGCCCGGCTCTGCTCTGACTCC	105420

QY	4918	cgagcagcagcagcagcttgccacacaaagggccaatctggaggaacaactctgtctggagggagaa	4977
Db	105419	CGAGCAAGGCGCGATGACACAAAGAGGCCAGTGAAGAGCAACTCTGTGTGAAGAGGAA	105360
QY	4978	aaaccaaaagatgagagagctctggccctgcctccacagccccggagagagagcctctgcag	5037
Db	105359	AAAGCAAAGATGAGAGAGGTGGGCGTGGCTGCTCCAGGCCCGGAGGGGAGGCGCTGGCA	105300
QY	5038	ccccagacaaggcacaagaacagccagcccaaccaactctagcagctattccaagcgg	5097
Db	105299	CCCCAGACAAGGGACACAGAAACAGCAGGGGCAACACAACTACACAGCATTTCCAAAGCG	105240
QY	5098	aagcgctcaactctggggccgggccaagaacaacactcttaacctgttaagggcgctgc	5157
Db	105229	AAAGCGCTCATCTCGGGGCGGGCCAGAAACACACTCTTCAACCTGTAAAGGGGGGTGCC	105180
QY	5158	aagcgagcagacaacaagaaagtctgcctccctggatctcccgagagcctaaatccgctc	5217
Db	105179	AAACGACACACACAGCAGAGAGGTGTGTGCCCTTGATCCCGAGAGCCTGAATATCCGGCTC	105120
QY	5218	aagtaattctctcttgcagaaggggtataggtctagaagaacggagaccoccgctctcaacc	5277
Db	105119	AAATTAATTCTCTTTCACAAAGGCGTGAAGGTCAAGAGCGGAGACCCCGCTTCTCAACC	105060
QY	5278	ttctgtcgggtgtagaagcgaagacgctctcaaccacatatgactgtgttcaactccct	5337
Db	105059	TTCTGTCGGGGTGAAGAGGAGAGCGCGTACACACATATGACTGTGTGCAATCCCT	105000
QY	5338	ggagatgctgcacaaagccccaacagaaagctctctctctgtgcctcttctccatctctg	5397
Db	104999	GGAGATGCGCCCAAGCCCAACAGAAAGCTTCTCTCTGCTGCTCTCTTCTTCATACCTCG	104940
QY	5398	ttctgttctctcttgcgtatgcagtcggggcctccctctgcgacaactcccttgtagctcaatc	5457
Db	104939	TTCTGCTTCTCTTGTGATGAGAGCGGGGCTCTCTTGCCACACTCTCTGAGAGCTTCATC	104880
QY	5458	ctcgagcgcgcggccctcctctgcctcctctctcatcagaatgcacttgggctgtgtctcc	5517
Db	104879	CTGACACCGCGGCGCTCTCTTGGCCCTCTCTCCACATCAATTTGGGGCGCTGTGGTTTTCC	104820
QY	5518	aagggcccttagaactcttctctgtttgttgctgctctgcgcaaaacccggagcaacttaag	5577
Db	104819	AAAGCCCTAGTAACCTCTCTCTCTGTGTGGCTCTGCGCAAAACCCGCAACTTCAG	104760
QY	5578	gaactctggggaaactctcttggccctcatcactctgaacaactgtctctcccaaaagaagcca	5637
Db	104759	GACCTTGGGGACCTCTGTGGGGCCTACTATACCTGAACACTGCTCCCAAAAAGAGCA	104700
QY	5638	aaactcaagagaaggtgtgcgcagagaaggaacctgtgaagagagctcgtctgcgcgttag	5697
Db	104699	AAACTCAAGGAAGAGTGGCGGCAAGAGCACTGTGAAGAGAGCTGTGCTGCGCTTAG	104640
QY	5698	agaaactcaaaaggtctcccgagttgaaactctgcgcaactctgcgcgcggagaaagccccag--t	5755
Db	104639	AGAACTCAAAAGGTCCCGAGTGTGAAGTGTGAAGTGTGCGCACTCTCCGGGAAGCCCCCAAGGCT	104580
QY	5756	gaacgcgcacgtgcacccgcgcgaagcagagggcccaactgcgacacaggtgcgcggggtctcc	5815
Db	104579	GACGCGCCAGCTGAACCCGGGCAAGAGAGGCCCACTGTGCAACAGTGCCTGGGGCTGTCC	104520
QY	5816	cggagagctctcagaagctctctctctgtctgtatgtgcgggaggaatgggggcgagagaccc	5875
Db	104519	CGAGAGCTCAAGTGTCTACTGTCTGTATGAACCGGAGGAAGTGGGGGAGAGGAGACCC	104460
QY	5876	ccaacccacaagaaggtctgcaaaactgtgtgcagcaagaagagctccgcgacagcccgcggg	5935
Db	104459	CCAACCGAAGAGGTGCGAACTAATAGTGAAGCAAGGAAGGTCCGGCAGAGCCCGGGCGG	104400
QY	5936	gaagcccaagagcactgtgtgtcatctgaagcctgtgcgtgttggaacggcggtctctactgt	5995
Db	104399	GAGGCCCAAGACACTGTAGTGCAATGAGCGTGTCTCGTGTGAACCGGAGCGCGTCTACTGTG	104340





Dh 9341 CACTOCTGACACAGAGTTGTCAATATGCGACAGCTGGTGGGAGAGTCAAGCGAG 9400  
Qy 301 ggggactatataagagctgcgcggggagagagagagagagagagagagagagag 360  
Dh 9401 GGGGACTATTAAAGAGTGGCGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 9460  
Qy 361 cttcttgagagaaag 420  
Dh 9461 CTTTCTGAGGCAAAAGAGAGTGGCGGGTGGAGTGGAGTGGAGTGGAGTGGAG 9520  
Qy 421 cactccctccctccctccctccctccctccctccctccctccctccctccctcc 480  
Dh 9521 CATCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 9580  
Qy 481 gcccgagtcagagctctcttgagagagagagagagagagagagagagagagag 540  
Dh 9581 GCCCGAGTCAGAGCTCTTTTGGAGAGAGTGGTTCCTCATGAGCAAAACAGAGTAC 9640  
Qy 541 cagcagagctcgcagagaaacatcagcgcctagagagagagagagagagagagag 600  
Dh 9641 CAGCAGAGCTCGAGAGAAACATCAGAGAGTGGAGAGTGGAGAGAGAGAGAGAG 9700  
Qy 601 ctaagctgcagacgc 660  
Dh 9701 CTAGCTGCGACGCGGAGCGGCTGCTGCGAGAGAGTATTAACCGGAGAGAGAGAG 9760  
Qy 661 agctatag 720  
Dh 9761 AGCTATAGAGAGAGTGGCGGCGAGAGAGTGGAGAGAGAGAGAGAGAGAGAGAG 9820  
Qy 721 taccaacag 780  
Dh 9821 TACCAACAG 9880  
Qy 781 cctgagcag 840  
Dh 9881 CCTGAGCTAG 9940  
Qy 841 cttcagagcttgag 900  
Dh 9941 CTTGAGAGCTTGAG 10000  
Qy 901 gtagcag 960  
Dh 10001 GTAGCAGAGTATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 10060  
Qy 961 gtagcag 1020  
Dh 10061 GTAGCAG 10120  
Qy 1021 ccgagcag 1080  
Dh 10121 CCGAGCCAGAT 10180  
Qy 1081 gctccctctgag 1140  
Dh 10181 GCTCCCTCTGAT 10240  
Qy 1141 acctcctcag 1200  
Dh 10241 ACCTCCTCAGAT 10299  
Qy 1201 agttcag 1260  
Dh 10300 AGTTGACAGAT 10359  
Qy 1261 gcccgag 1320  
Dh 10360 GCCCGAT 10419  
Qy 1321 cagcag 1380  
Dh 10420 CAGCAGAT 10479

Qy 1381 catgcag 1440  
Dh 10480 CATGCCAGAT 10539  
Qy 1441 gtagcag 1500  
Dh 10540 GTAGCAGAGTATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 10599  
Qy 1501 ttagcag 1560  
Dh 10600 TTGAGCCAGAT 10659  
Qy 1561 tccag 1620  
Dh 10660 TCCAGAGAGTATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 10719  
Qy 1621 agcag 1680  
Dh 10720 AGCAGAT 10779  
Qy 1681 cctccag 1740  
Dh 10780 CCTCCAGAT 10839  
Qy 1741 cag 1800  
Dh 10840 CAGAT 10898  
Qy 1801 ctgag 1860  
Dh 10899 CTGAT 10958  
Qy 1861 gctgagcag 1920  
Dh 10959 GCTGTGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 11018  
Qy 1921 agcag 1980  
Dh 11019 AGCAGAT 11078  
Qy 1981 ttagcag 2040  
Dh 11079 TTAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 11138  
Qy 2041 ctgag 2100  
Dh 11139 CTGAT 11198  
Qy 2101 agcctgag 2160  
Dh 11199 AGCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 11258  
Qy 2161 gtagcag 2220  
Dh 11259 GTAGCAGAGTATGATGAGAGAGAGAGAGAGAGAGAGAGAT 11318  
Qy 2221 agctgag 2280  
Dh 11319 AGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 11378  
Qy 2281 ctgag 2340  
Dh 11379 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 11438  
Qy 2341 atgag 2400  
Dh 11439 ATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 11498  
Qy 2401 aagcag 2460  
Dh 11499 AAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 11558

QY 2461 tggcccgccgtgggagccggagcccgcccgactccctgagctggacaaagggcgcc 2520  
Dd 11559 tggcccgccgtgggagccggagcccgcccgactccctgagctggacaaagggcgcc 11618  
QY 2521 aatgccagagcttaagcccgagctgtttgaagacccttcctggccttcagctact 2580  
Dd 11619 aatgccagagcttaagcccgagctgtttgaagacccttcctggccttcagctact 11678  
QY 2581 gaccccaaaagacaactggctcctcctccttgggtacaaagcccccctgggttcc 2640  
Dd 11679 gaccccaaaagacaactggctcctcctccttgggtacaaagcccccctgggttcc 11738  
QY 2641 gctccagaccccaactacagagcttttgaagcttttcccgacacacacccgtcagatca 2700  
Dd 11739 gctccagaccccaactacagagcttttgaagcttttcccgacacacacccgtcagatca 11798  
QY 2701 gggagagcgcccaaccccttggctggcagaggaacacccctgggtatccttgcacag 2760  
Dd 11799 gggagagcgcccaaccccttggctggcagaggaacacccctgggtatccttgcacag 11858  
QY 2761 tgggattgacccctggcagcttaccagggcctggagcaaggtgggaagcctcagat 2820  
Dd 11859 tgggattgacccctggcagcttaccagggcctggagcaaggtgggaagcctcagat 11918  
QY 2821 gcatcagcaaaagggacacccatgaagcttggcctcctggccttccagagagagac 2880  
Dd 11919 gcatcagcaaaagggacacccatgaagcttggcctcctggccttccagagagagac 11978  
QY 2881 ccccttgggagaggtggcctccttggcgggagcttccagagagaggtgggtggg 2940  
Dd 11979 ccccttgggagaggtggcctccttggcgggagcttccagagagaggtgggtggg 12038  
QY 2941 gtagagagagagaggtggcctcctgagtgcccgaggtggccaggtgacccgtgg 3000  
Dd 12039 gtagagagagagaggtggcctcctgagtgcccgaggtggccaggtgacccgtgg 12098  
QY 3001 ctgagagagagagaggtgttccagccagcttggccttccagagagagagagag 3060  
Dd 12099 ctgagagagagagaggtgttccagccagcttggccttccagagagagagagag 12158  
QY 3061 cccacagagagagagagagctgagagagagagagagagagagagagagagagag 3120  
Dd 12159 cccacagagagagagagagctgagagagagagagagagagagagagagagagag 12218  
QY 3121 ctggcag 3180  
Dd 12219 ctggcag 12278  
QY 3181 atctcag 3240  
Dd 12279 atctcag 12338  
QY 3241 tggccacctcagggaggtccgtcctcctcctcctcctcctcctcctcctcctcctc 3300  
Dd 12339 tggccacctcagggaggtccgtcctcctcctcctcctcctcctcctcctcctcctc 12398  
QY 3301 gtcag 3360  
Dd 12399 gtcag 12458  
QY 3361 gggag 3420  
Dd 12459 gggag 12518  
QY 3421 aagcctgtgtggcag 3480  
Dd 12519 aagcctgtgtggcag 12578  
QY 3481 ttacgag 3540  
Dd 12579 ttacgag 12638  
QY 3541 cagtggtgccccaaag 3600







Db 2189 AGGCTACAGGCAAGTCGCGGCTGACAGCTTCTCCAAAGTTGCTGCGGGGTGAGCGGAC 2248  
QY 2269 TATCGCGGCTGTGTCTCAAGCGCCCTGCGCAAGAGAAAGTGTGCTCCGATCCTGCGG 2328  
Db 2249 TGTCCGCGCTGTGTCTCAAGCGCCCTGCGCAAGAGAAAGTGTGCTCCGATCCTGCGG 2308  
QY 2329 CTGCGAGAGAGCTCGTGTGAGAGAGCGCAAGAGTGTGCTGAGAGAGAGAGAGAGAG 2388  
Db 2309 CTGCGAGAGAGCTCGTGTGAGAGAGCGCAAGAGTGTGAGAGAGAGAGAGAGAG 2368  
QY 2389 AAGGACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2448  
Db 2369 AAGGACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2428  
QY 2449 GCGCAAGAGT 2508  
Db 2429 GCGCAAGAGT 2488  
QY 2509 GCGCAAGAGT 2568  
Db 2489 GCGCAAGAGT 2548  
QY 2569 TCGCTACGCTGT 2628  
Db 2549 TCGCTACGCTGT 2608  
QY 2629 CTTGGGAGT 2688  
Db 2609 CTTGGGAGT 2668  
QY 2689 GCTGCAAGT 2748  
Db 2669 GCTGCAAGT 2728  
QY 2749 GCTGCAAGT 2808  
Db 2729 GCTGCAAGT 2788  
QY 2809 AAGGCTCAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 2868  
Db 2789 AAGGCTCAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 2848  
QY 2869 CAGGAG 2928  
Db 2849 CAGGAG 2908  
QY 2929 GAGGAG 2988  
Db 2909 GAGGAG 2968  
QY 2989 GCTGCAAGT 3048  
Db 2969 GCTGCAAGT 3028  
QY 3049 CCACTGT 3108  
Db 3029 CCACTGT 3088  
QY 3109 CTAATGT 3168  
Db 3089 CTAATGT 3148  
QY 3169 AAGGAG 3228  
Db 3149 AAGGAG 3208  
QY 3229 TGGGAG 3288  
Db 3209 TGGGAG 3268  
QY 3289 ACCGAG 3348  
Db 3269 ACCGAG 3328

QY 3349 GAGGAG 3408  
Db 3329 GAGGAG 3388  
QY 3409 CAGGAG 3468  
Db 3389 CAGGAG 3448  
QY 3469 CAGGAG 3528  
Db 3449 CAGGAG 3508  
QY 3529 CAGGAG 3588  
Db 3509 CAGGAG 3568  
QY 3589 CAGGAG 3648  
Db 3569 CAGGAG 3628  
QY 3649 CAGGAG 3708  
Db 3629 CAGGAG 3688  
QY 3709 CAGGAG 3768  
Db 3689 CAGGAG 3748  
QY 3769 CAGGAG 3828  
Db 3749 CAGGAG 3808  
QY 3829 CAGGAG 3888  
Db 3809 CAGGAG 3868  
QY 3889 CAGGAG 3948  
Db 3869 CAGGAG 3928  
QY 3949 CAGGAG 4008  
Db 3929 CAGGAG 3988  
QY 4009 CAGGAG 4068  
Db 3989 CAGGAG 4048  
QY 4069 CAGGAG 4128  
Db 4049 CAGGAG 4108  
QY 4129 CAGGAG 4188  
Db 4109 CAGGAG 4168  
QY 4189 CAGGAG 4248  
Db 4169 CAGGAG 4228  
QY 4249 CAGGAG 4308  
Db 4229 CAGGAG 4288  
QY 4308 CAGGAG 4368  
Db 4289 CAGGAG 4348  
QY 4368 CAGGAG 4428  
Db 4349 CAGGAG 4408



Db 241 GAGCAGAGAGCCCTGCCCCACAGACAGAGGCTTCAGAGGAGGCGCGCTTCCCTGACTAC 300  
QY 790 ggcctccagagacagcagccctaccagcagcgtatgctggtgagagagccttaagct 849  
Db 301 GGGTCACAGACAGACAGCCCTTACCCAGGCGCTTAAGCTGGAGAGAGACCTTCAAGCT 360  
QY 850 tggggggcccaagcagccacaccccaagccgcaagcactaccctgcaagggtggccaag 909  
Db 362 TGGGGGGCCCCAGACGACACACCCCAAGCCGACACCACTACCTGCAAGGAGGCGCAAG 420  
QY 910 tatgatgagaaacttgatgaaagaagcagctgccccagcagcagctatgacagcag 969  
Db 421 TATGATGAGAACTTGATGAAAGAAAGACAGAGTGCCTCCCAAGAGAGTATGCAAGAGAG 480  
QY 970 ggcgcagagtgccctcttgagactacacccctgacgttcagagcagcagcagcagcag 1029  
Db 481 GGGCGCCAGAGTGGCCCTTTCGAGCTACCTCCCTGACAGTCCAGAGCCACCGCGGCCAG 540  
QY 1030 cagccctcgtgcataccacagctccaaagcagagctgagcagaaacgaattgcctccct 1089  
Db 541 CAGCCCTGCGCATACCCCAAGCTCCAAAGGACAGAAAGCTGCAAGACGACATTCCTCCCT 600  
QY 1090 ctgcccctcccccagaggtaccacattccagcatccagctcctcccccacccctcc 1149  
Db 601 CTGCCCTTCCCCAG----- 615  
QY 1150 acctactcctctctgcaagggtggtgagagggggccacactcctataagagtgcaca 1209  
Db 616 -----GGGTGTGGGAGAGGGG--CCACTCCTTAAGAGATTGCACA 654  
QY 1210 gcaacagactgcccagcccaatgaagcgccttactgcaagctcaacccctggcccgagg 1269  
Db 655 GCACGACTGCCCCAGCCCCATACAGGCGCTGTACTGCAAGCTCCAGCCCTGGGCCGGGG 714  
QY 1270 cagcgggtccagaaacttcaatgctcctacagctcgggccgctcctacgtatgaacagag 1329  
Db 715 CAGCGGGTCCAGAACTCTCATCTACAGAGTGGGGCGGCGCTCAGCTATGACAGACAG 774  
QY 1330 cagcagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1389  
Db 775 CAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 834  
QY 1390 gaaacccctccattcaaaaaactcgcagaatcagcactacagcagcagcagcagcagcag 1449  
Db 835 GAAACCTTCATACAAAACCTCCGCAAGTATACAGATAGGGGAGCAAGGCCAGGGC 894  
QY 1450 tactcagcagcagcagcagcagctccgagcccccagcagctactacagagactccagccc 1509  
Db 895 TACTGCGACCGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 954  
QY 1510 agctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1569  
Db 955 AGCTCCAGCCACTACCCCGCGCTCCGTGGGCGCTCCCTCTCTCAAGTTCACACCG 1014  
QY 1570 tgcgcgtgatgacaaactggaagaaacttccctacagcagcagcagcagcagcagcagcag 1629  
Db 1015 TCGCGCGTGAAGCCAAACCTGGAAGCTTCCCTACAGCAGCAGCAGCAGCAGCAGCAGCAG 1074  
QY 1630 gscctcccccagagagatcaatgacacagccacttatagccccctgctcaatcccccaca 1689  
Db 1075 GCGTTCGCCGAGGAGATCACTACACAGCCACTTATCCCTGCTCAATCCCTCCCA 1134  
QY 1690 acgagatgcacacagctctgtgagacacccagctgagcaactgaaagccctcagaagagac 1749  
Db 1135 ACGAGTGCACACAGCTCT----- 1152  
QY 1750 aagctccctgagaaactgctgctgagatcctcaagctgcaagcctcagcagcgtgacctta 1809  
Db 1153 -----CCTCAGAG--CCTCAGAGGCGGCTGACCTCA 1179  
QY 1810 caggtgagaaactcctcaacacagctccagcagcagctgctcctcctcaagagctcgtgagc 1869  
Db 1180 CAGGTGAGAACTCTCCAAACGCTCCAGAGCTCTGCTCTCCAAAGGCTGCTGTGGCG 1239

QY 1870 cagaagaaggtgtcagaagactggtccagagaccccaagagcagcagcagcagcagcagcagc 1929  
Db 1240 CAAAGAAAGAGGTGCAAGAACTCTGCTCCAGAGACCCCAAGAGAGCACTAAAGCCAGAC 1299  
QY 1930 tgcagcccccgaagagagcagctactcagccagcccgcaagcagcagcagcagcagcagcagc 1989  
Db 1300 TGCAGCCCGGAAGGAGGAGCGCTACTACAGCCGAGCCCGAGCAGACACCGCTGCAAGCCG 1359  
QY 1990 ccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2049  
Db 1360 CCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1419  
QY 2050 tccgagagcccaactgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2109  
Db 1420 TCCGAGAGACCCAGCAGGAGCGAGCTTCCTCTACTGCAAGAGCCCTGAGCCCTGCC 1479  
QY 2110 aaggtcaacagcaactcgaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2169  
Db 1480 AGGCTACACAACTCGAAGGCCAAGCCAGAGTCCGTGTCCACTTCTGTGACCTCT 1539  
QY 2170 cctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2229  
Db 1540 CCGAGCAGATGTCACAAATCTGACAGCTCTTCCAGAGCTTACAGGCACTGCTGCCG 1599  
QY 2230 ctgagacactcctccaaagtctgctgagcagcagcagcagcagcagcagcagcagcagcagc 2289  
Db 1600 CTCGAGAGCTTCTCAAGTTCGTGGCGGAGGAGCAGGAGACGTCCGCGCTGCTGCTCAGC 1659  
QY 2280 ggccttgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2349  
Db 1660 GCGCTGACAGAGAGAGCTGAGCTCCAGAGTCTGAGGCTGAGAGAGCAGCAGTGTAG 1719  
QY 2350 aagcgcgaagaagcttgggtgaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2409  
Db 1720 AAGCGAGCAAAAGCTTGGCTGAAGCAGCCAGCTGTGTAAAGAGACAGCAGCAGCAGCAGC 1779  
QY 2410 ttctgctgagagacacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2469  
Db 1780 TTCTGCTGGAGAAACACAGCCCGCTGCTGAGCTGTGTGAGCAGAGTGTGCGGCGCGG 1839  
QY 2470 cctggagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2529  
Db 1840 CCGAGGAGAGCGAGAGCGCTGCCAGACCTCTGAGCTGTGAGAGAGGCGGAGCTGAGCAG 1899  
QY 2530 gactcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2589  
Db 1900 GACTTCAGCCAGGGGCTGTTGAAGACCTTCGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTG 1959  
QY 2590 aagc 2649  
Db 1960 AAGACACTGGTCT 2019  
QY 2650 cccactacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2709  
Db 2020 CCGACTACAGAGAGCTTGTGAGCTTCCCGAGACAAACGCTGACACTAGAGGAGCAGC 2079  
QY 2710 gccacccctctgctgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2769  
Db 2080 GCCAACCCCTTGTGCTGCGCAGAGGAAACCTGGGGAGAGCTGTGCCAAGGTGGGAGATTG 2139  
QY 2770 caccctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2829  
Db 2140 CAGCCGAGGAGCTTACAAAGGCGCTGAGAGAGGCTGAGAGAGGCTTACAGTGGCATACAGC 2199  
QY 2830 aaagggagaccccaatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2889  
Db 2200 AAAGGGAGACACCACTGAGAGCTTCCGCTGCTGAGCTTCCAGAGAGAGAGCCCTCTGGG 2259  
QY 2890 gaagaagctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2949  
Db 2260 GAAAGAGTGGCTGTGCTGCGGAGACTTACAGCAGAGAGAGAGTGGGTGTAAGAG 2319

QY 2950 gagcaggtgggtgctgctgagtgcccgagtggtggaaggtggaaggtggtggtgagac 3009  
DB 2320 GAGGAGGTGGGTGCTGCTGAGTGGCCCGAGGTGGCCAGAGGTGACCGGTGGTGGAGAC 2379  
QY 3010 agcggcaactgctgttccacacgcagcagcttggggacccctccactgtgccaacacagc 3069  
DB 2380 AGCCGGGACACTGCTGTTCACCGCGGACTTCGGGGGACCTCCCACTGCTGCCACCCAGC 2439  
QY 3070 aggaagagagacactgaaagctgagagagagtagctccctctatgtagctctggcagc 3129  
DB 2440 AGGAAGGAGACCTGGAAGCTGAGAGAGAGTACTCTCTCTATGTAGAGTCTTGGGAGC 2499  
QY 3130 ccgagcagaggtgctgctgagcagcagcagcagcagcagcagcagcagcagcagcagc 3189  
DB 2500 CCCGAGCAGAGGCTGGCATGCAGAGCCCGCTGTCCCAAGGCCCACTCATCTGCAAC 2559  
QY 3190 aagagagagtgagaggtgctgagcagcagcagcagcagcagcagcagcagcagcagc 3249  
DB 2560 AAGAGAGAGTGGAGAGGTGCTGGATCTCCAGAGCGGCTGGGGGCTCTCCGTGCCACCTC 2619  
QY 3250 tcagggagtcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 3309  
DB 2620 TCAGGGAGTCCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2679  
QY 3310 tggttgagtcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 3369  
DB 2680 TGGTTGAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2739  
QY 3370 gctcagggattccacactcagcagcagcagcagcagcagcagcagcagcagcagcagc 3429  
DB 2740 GCTCAGGGGNTCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2799  
QY 3430 gtgcccagggggccatcgcaaaagagagcctgtgcaaggggggcaaaagcttaccagc 3489  
DB 2800 GTGCCGAGGCGCCATCGCAAGAAAGAGCCTGTGCCAGGGGCAAAAGCTTACGGAGC 2859  
QY 3490 cgtcgggtgacagggggtgcccagagcagagagcagcagcagcagcagcagcagcagc 3549  
DB 2860 GCTGGGTGTCACCGGGGGTGCCTGAGCGCGAGAGTCTCCATGACAGGCGCACGAGTGC 2919  
QY 3550 cccaaagaccttctgctcctgaaatcgtgcaaggggggggggggggggggggggggggg 3609  
DB 2920 CCAAGAACCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2979  
QY 3610 gctgagccacagcgggggggctgagaggggctgagaggggctgagaggggctgagaggg 3669  
DB 2980 GCTGAGGCCCGAGCGGGGGGCTCTGGAAGGGGCTCCCGAGGATGTACTGCTCTCTC 3039  
QY 3670 acggtcctgagtgagcccgacagcccgagcccgagcccgagcccgagcccgagcccg 3729  
DB 3040 AGGGCTGTAGTGAGCCCGGACCGCGCGGACCGCGCGGACCGCGCGGACCGCGCGGAC 3099  
QY 3730 ccagacaaactggggggaagcagcagcagcagcagcagcagcagcagcagcagcagcag 3789  
DB 3100 CCAGACAACTGGGGGGCAAGCAGCAGCGCGCTTCAAGTGGGGCAGCGGTGGGAAG 3159  
QY 3790 cctcacaacagcgtcctcagccagcagcagcagcagcagcagcagcagcagcagcagc 3849  
DB 3160 CCTCACCAAGGCTGCTTCCAGCGCCGAGCAAGCGGGGCTGCTGCTGCTGCTGCTGCT 3219  
QY 3850 agcagcccgatggggtccaaagcagcagcagcagcagcagcagcagcagcagcagcagc 3909  
DB 3220 AGCAGCCGATGGGCTCCAAAGCAAGAGAGAGAGACTCACCCAGCAGCGCTGGCAGGAC 3279  
QY 3910 cagcgtccatgatacttgggtcagcagcagcagcagcagcagcagcagcagcagcagcag 3969  
DB 3280 CAGCGCTCATGATCTTGGGTGACGACCAAAACCAAGAGATCTTCCACTCCCAAGCGG 3339  
QY 3970 cagagccctcagggcgctcccaactcctgctgcaacagagcagcagcagcagcagcagc 4029  
DB 3340 CGAGGGCCCTGTGAGGGCGGCTCCCTCACTGCGGCTGCGACCAAGAGAGTCTCGACAC 3399  
QY 4030 agccacttgcgcacattcaaggtctccagcagcccccagaggggaggggtgagc 4089

DB 3400 ACCCACTTGGCCCGCACATTCAAGGTCTCCAGACGCCCCAGAGAGGGCAGGGTGAC 3459  
QY 4090 caggggcaaggggtcccccacaaactgtgagcagcagcagcagcagcagcagcagcagc 4149  
DB 3460 CAGCGGGCAAGGGTCCCAAAAGCTGTGAGGAGCAAGAGCTCTGACCGGCCCTCCAT 3519  
QY 4150 ggcctcacaaggaagtcggccttcagcagcagcagcagcagcagcagcagcagcagcagc 4209  
DB 3520 GGGCTCAAAAGGAAGTCTGGCTTCTATGGCGCGCTCCCAACAGAGAGGAACTGGTC 3579  
QY 4210 ttggggcagcggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4268  
DB 3580 TTGCGGAGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 3639  
QY 4269 ggaagcctgaggttcccccacaaacttccagagagatgtcttcccaagaagcagcag 4328  
DB 3640 GAGAGGCTGAGGGTTCCTCCCAACCTCTTCAAGAGGATGTCTTCTCCCAAGAAAGCAAG 3699  
QY 4329 cccacaagggcaatggcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4387  
DB 3700 CCCACCAAGGGCAATGGCGAGCGCTCCACAAAGCTCCCAACCGCGGAGACCCCGATGCC 3759  
QY 4388 tgcctcaagctgcctctc-ggcagccttccaggggggcatgaagaccaggtgctgcca 4446  
DB 3760 TGCCTCAAGCTCGCTCTCGGCGAGCTTCCAGGGGGCCATGAGACCAAGGTGCTGCA 3819  
QY 4447 ccccggaagggcggggctgaagctggaagcagcagcagcagcagcagcagcagcagcagc 4506  
DB 3820 CCCCGAAGGGCGGGGCTGAGCTGGAAGCCATCTGTGAGAGATACCTTCGCCGACG 3879  
QY 4507 ctcaagaagttgcctgtaagggcgaggggctctctctgtaactctctgagccatcc 4566  
DB 3880 CTCAAGAGTTCATGTAAAGCGCCAGGGGCTCTCTGTAATCTCTGAGCCCATCC 3939  
QY 4567 ctttcccaaaagacgttggtcagaggtgctgagggcagcagcagcagcagcagcagcagc 4626  
DB 3940 CTTTCCGACAAAGACCGTGGGCTCAAGGGTCTGAGGGGCGCCAGCCAGTGGGGTGAAGAA 3999  
QY 4627 ggcctggtaagtgggcacggcgagagctcccaactcttggggctgctcgttatgc 4686  
DB 4000 GGCCTGGTAAATGTGGGCAACCGGCGAGAGCTCCCAACTTCTGGGGCTGATCCGTTATGC 4059  
QY 4687 agaatccaaacaaagatccttaaaaggcaactcatgaacagtaagaactgtctct 4746  
DB 4060 AGAAATCAACCAACAGATCTTAAAGGCAAACTCATGA----- 4099  
QY 4747 actgactgtttcaaaacagagccttcacatcccgaggccctgagcctgggggact 4806  
DB 4100 ----- 4099  
QY 4807 ggcctggcgtcagaagagagcgggaaaggccgggaggggggggggggggggggggggg 4866  
DB 4100 ----- 4099  
QY 4867 ggcctggcgtgagagagcggccttatcttggccggctctgctcctgactcccgagcagc 4926  
DB 4100 -----ACAGG 4104  
QY 4827 ggcctggcgtgagagagcggccttatcttggccggctctgctcctgactcccgagcagc 4986  
DB 4105 GGCAGTGGCACACAGAGGGCCAGTGGAGCAACTCTGTGTGGAGGAGGAGAGCAAG 4164  
QY 4987 atggagagcgtgggctcctccacagcccgaggggaggggggggggggggggggggggg 5046  
DB 4165 ATGAGGAGGTGGGCTGCTCCAGCCCCCGAGGGGCGGCTTCCAGCCCGCCAGACA 4224  
QY 5047 agggcacaagaaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5106  
DB 4225 AGGGCACAGAAACAGCCAGCCCAACCACTACAGCATCTATTCAGGCGGAGGCGCTC 4284  
QY 5107 actcggggcggggccaagaaacacacactcttccactgtaaggggggtgccaagcgagc 5166



```

Db      4285  ACTGGGGCCGGGGCCAAAGAACACCACTCTTACCTCTGTAAGAGGAGGCGTGCACAGCAGCA 4344
Qy      5167  cgaacaagcagatgctgccccgagatcccgagagagccttaactcgcctcaataatt 5226
Db      4345  CGACAGCAGAGTGTGCCCCCGGATCCCGAGAGGCTTAATCCGCTCAATGATCAT 4404
Qy      5227  tcccttgcagacgagctgagtgtagacagcggccccgctctcaaccttgctcg 5286
Db      4405  TCCTCTTGCAAGCGCTGAGTGAAGACAGCGAGCCGCCCTCTTCAACCTTGCTGCG 4464
Qy      5287  gtggaagacgagacgcttaccacacatgacactgtgtcaactccctcgagatg 5346
Db      4465  GTGAGAGAGCAGAGAGCGCTTACACACATGACACTGTGTCACTCCCGCGAGAGAG 4524
Qy      5347  ccaagaccccaaggaagaccttccctctgctcctcctcctcctcctcctcct 5406
Db      4525  CCGAGCCCGACAGAGAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4584
Qy      5407  tcccttgcagacgagcggcctccctcgagacacacacacacacacacacacac 5466
Db      4585  TCCTTGATGACAGCGGCGGCTCCCTGCGCACACTCCCTGGAAGCTCCATCTGAC 4644
Qy      5467  cggccctcttgcctccctctcctcctcctcctcctcctcctcctcctcctc 5526
Db      4645  CGGCGCTCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4704
Qy      5527  agtaactctgcttctgttctgctcctcctcctcctcctcctcctcctcctc 5586
Db      4705  AGTACCTCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4764
Qy      5587  gacctctgtgagccttactactcctgagacacacacacacacacacacacac 5646
Db      4765  GACCTCTGTGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4824
Qy      5647  gagaaggtgagcagcagagacacacacacacacacacacacacacacacac 5706
Db      4825  GAGAAGGTGAGCAGAGACACACACACACACACACACACACACACACACACAC 4884
Qy      5707  aaaggtccagatgtgacagctgcgcacacacacacacacacacacacacac 5764
Db      4885  AAAAGGTCCAGAGTGTGACAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4944
Qy      5765  gctgaccccgagcagcagcagcagcagcagcagcagcagcagcagcagcag 5824
Db      4945  GCTGACCCCGAGCAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5004
Qy      5825  cagagctgctactgctgtgagtgagcagcagcagcagcagcagcagcagcag 5884
Db      5005  CAGAGCTGCTACTGCTGTGAGTGCAGGAGGAGTGGGAGGAGGAGGAGGAGGAG 5064
Qy      5885  aaggggtcgaacaacatgagtgacgaagagagcctccgacagaccccgag 5944
Db      5065  AAGGG----- 5069
Qy      5945  gagcactgagtgatgagagcctgtgacacacacacacacacacacacacac 6004
Db      5070  --ACACTGAGTGTGATGAGAGCTGTGCGCGTGTGAGACCGAGCGCTCTACTG 5127
Qy      6005  aaagctcttggagtgag 6022
Db      5128  AAGCTCTTGGGCTGAG 5145

```

```

RESULT 6
LOCUS      HSA271790      5667 bp      mRNA      PRI      17-JUN-2001
DEFINITION Homo sapiens mRNA for retinoid-acid induced protein 1 (Rai1 gene).
ACCESSION  AJ271790
VERSION     AJ271790.1 GI:12053792
KEYWORDS   Rai1 gene; retinoid-acid induced protein 1.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE  1 (bases 1 to 5667)
AUTHORS   Seranski P., Hoff C., Radelof U., Henry S., Reinhardt R.,
            Schwartz C.E., Heiss N.S. and Poustka A.
TITLE      Rai1 is a novel polyglutamine encoding gene that is deleted in
            Smith-Magenis syndrome patients
JOURNAL    Gene 270 (1-2), 69-76 (2001)
MEDLINE    21297181
REFERENCE  2 (bases 1 to 5667)
AUTHORS   Seranski P.
TITLE      Direct Submission
JOURNAL    Submitted (31-JAN-2000) Seranski P., Molecular Genome Analysis,
            Deutsches Krebsforschungszentrum Heidelberg, Im Neuenheimer Feld
            280, 69120 Heidelberg, GERMANY
FEATURES
Source     Location/Qualifiers
            1..5667
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="17"
            /map="17p11.2"
            1..5589
            /gene="Rai1"
            1..5589
            /gene="Rai1"
            /function="unknown"
            /product="retinoid-acid induced protein 1"
            /protein_id="Cac20423.1"
            /db_xref="GI:12053793"
            /translation="MOSFRFGKFGKQKNTQSTSELENTROSDAGLSQDR
            ILAKIYVNPQYPSYBEGAGTSSGTAAYADKHGSAALPTQOGIAGPAPGIV
            QDSPPYRIVAGESLSLAWGAPQPPQPPQPLAGAGAKIDENLKRAVPSQKQED
            GAVPEKTHSLHVOQPPQPLAVKIQKQKQNDIASPLPFGWAGGSHYKSC
            APTAPQHPRLTASSSLAPQOEVNLIAYOGSRLSDQCCQCCQCCQCCQCCQCC
            AQTLEIYONLAKTOYGOOGYQCPDAVTEPQYTOYFSPSSSHSPASVGRSPSY
            SNRSPIMNLEMPYHQDPLSTGAFAGTIDSHHMLNPSNPDATSSRADTALT
            SYENISNTVOQLLSKAAIPQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK
            SEPSSTPOSTHAPFORADLSDSEPLERSLFKQAGSPARNSKAKESYST
            CVSTSPDMKSRSDSFSQSLHSPLDSFSTFVAGRDQPRLLSALAQDLSELG
            LQALIERADKAWAEALSLKDSPLPFLNHSACIDVAKSAPRGPPEALPDSL
            QLDKGNANQDSPLFEDPVAFAFPDPPKKTGGLSFGRPTLGPADPPTAFCF
            PDRTASSADSNAPFAPEENLGDACPRWGHQBELTKLEGGQASDGLSKDTHEA
            SATGROEDPGRKVALSLGDPKQOEGYKREHAGGLIGDEYAKADRLDERSHC
            SLDRDPLPLPTSKEDIELEAEYSSLCILSPQKQPLQMDLSPKAPLICTREA
            VEEVLDSKAGWSPECHLSESVTLGPTVGESEVSNWESSLSHMKPGEEDGGERA
            PGGSTTSDSLAOKPNPVAPEALPAKPEPRKSLRERVHRLDEADSCRAV
            LPKDLIPSECTGPPQGMGAGAPGASBGLPRMCTSLTALSERPTGPGILTTT
            PAPPDLGKORAAFKSGKVGKSPKASSPALPVASDSPPGSKTEKTDSPS
            TPBKDORSWILASRTKQOETFHSKRRRSPRLPNCRAKTLIDNSHLPAHFVSSP
            QKGRYSQARVYRPGAKSLDRLPLHLKKAFTAPVPTKRLVTRSSSSSNA
            SGNQDGRKEPESPTLFRMSPEKAKPKKNGEAPKLPETPDACLKLASBA
            FQAMNTKTLVPRKRGKLEALVOKITSPILKFKAKAPASPNPISLSDNDRG
            LKAGGSPVGEGLVWVGQKLPFGADPLCNPFLNSLGLKLMNAGTQGASD
            NSGGGKKRKMKEGLASOPPEGRPOQTAOKQGHNTSYSKRRRLRGAANT
            TSSPCGKAKRROOVPLDPAPEPILKLTISCKRLSDRSPAPSPVPRKEDA
            FTTCIVYNSPDAPKHKRPTSSASSSSSSSSTDAAGALATLRGGSLPRESL
            PLSTVHLTPVYSKALSTCVCCQNPANFKLDGLGPTTPEHCLPKKKRLKEL
            VREGICEASLPLERTLKGPECAALATAGPDPGDPKQPLSTAKVLSRL
            QSCYCCDGEDGEAEAPADKGMVHECAVWTGVLVAGLFLQEMKAVVDHG
            GYVALAPGPESLPGIRFASLFOGSPWCDCPLTASPSMSGCAAKKXGPLGAT
            KQASPTTTRVAPMOVARRGCTGKNSPKSROTLRPHSSNPPLISLSPPLIQ
            OFTPSVCLDLCWAMPQTM"
BASE COUNT 1181 a 1962 c 1664 g 860 t
ORIGIN
Query Match 76.8%; Score 4627.6; DB 9; Length 5667;
Best Local Similarity 92.7%; Freq. No. 0;
Matches 511; Conservative 0; Mismatches 9; Indels 396; Gaps 10;
Qy      490  atgcagctcttgcagaaagtggttccatgcgaacaacagactaccagagacc 549

```





QY	2710	graaacccctcttcgagccagagaaacccgaggagatctgtccagatgtagattg	2765
Db	2080	GCCAACCCCTTTGCTCGCGCCAGAGAAACCTGAGGGGATCTTGTCCAGAGTGGGAAATTG	2139
QY	2770	caacccctcgagagcttaccaaagagccctcgagagctgtagaagagctccagatgatacag	2829
Db	2140	CACCTTGGCGAGCTTACCAAGGCGTGTAGAGAGGCTGGAAAGGCTCCAGATAGCGATCCAGC	2199
QY	2830	aaagggagacacccaatgaggtcttgccctcgtcgtcttcagagagagagaccctctgg	2889
Db	2200	AAAGGGAGCACCCATAGAGCTTTCGGCTCCGCTCCAGGGCTTCCAGAGAGAGAACCCCTCGGG	2259
QY	2890	agaaaggtgagccctcgttcccccggagactccaagcagagagagaggtggtgtgtgtagagag	2949
Db	2260	GAGAGAGTGGGCTCTGTGTCCGGGGAGATTCAAGCAGAGAGAGAGTGGGTGTGAAGAGAG	2319
QY	2950	gagcgaggtggtgtcgtcgtcagttgcccgaggtgtgccaaaggtctgacccggttggtgtgagag	3009
Db	2320	GAGGAGAGTGTGGGTGTGTGTGATGTCCCGAGAGTGGACAAAGGCTTATGACCTCTTGGAGAGAC	2379
QY	3010	agccggagacctgtcgttccacccggacagatcttcggaggacctccaatgtgtccacccacagc	3069
Db	2380	AGCCGGGACCTGTCTTCCACCGCCGACCTTCGGGACCTCCCATGTCTGTCCACCCACAGC	2439
QY	3070	agagaaagagagaccttgaaagactgtgaagagagagatcctccctcatgtgaagctctctggagag	3129
Db	2440	AGAGAAAGAGACCTGTGAGAGCTGAGAGAGAGAGATCTCTCTCCATGTGATCCTCTTGGAGAC	2499
QY	3130	cccgagagagagagccctgtgagatcagagagcccgctgttaccacaaagagccacataatgtacac	3189
Db	2500	CCCGAGAGAGAGCCCTGTGATGTGAGAGAGACCGGCTGTATCCAGAGGCCCATATGTATGAGAC	2559
QY	3190	aagagagagaggttgagagagaggtgtgtgtgaatccaaagagccggtctctccgttgagacctc	3249
Db	2560	AAGAGAGAGAGGTGAGAGAGAGGTCTGTGATCCAAAGGCCGGCTGTGGGCTGTCTGTGCACCTTC	2619
QY	3250	taagggagagctccgttcatcctcgtctggccctaaagtgtggacccgagtccaaggttccagagct	3309
Db	2620	TGAGGGAGAGTCCGTCTACTCTGTGGGCGCTTAAAGTGGGACCCAGTCAAAAGTGTCCAGAGAC	2679
QY	3310	tgggttgagtcctcctctgttcacacatgaaagcagaggtgaagaggtggtctgagtgtggagag	3369
Db	2680	TGGTTTGAAGTCTCTCTGTGACACATGTAAAGGCCAGGTGAAGAGGGGCTGTGATGGAGAGACA	2739
QY	3370	gtctccaggtggtatccacacacactgtgagagcctctctgtgcccagagagccacaagctgct	3429
Db	2740	GCTCCAGAGGGGATTTCCACACCTCCGAGAGCGCTTCTGTGGGCCCAAGAGCCCAAAACCTGTGCT	2799
QY	3430	gtgccagagagccgacatctgacaaagaaagagagcctgtgcccagagggagaaagattacagagac	3489
Db	2800	GTGCCAGAGGGGCCCATCCCAAAAGAGGCTGTGTGCCAAGGGGCAAAAGCTTATCGGAGAC	2859
QY	3490	cgctcggtgtgacccggtgggtgtgtcccgagaggtacgttcccatgtgaggtgacacaggtgtgt	3549
Db	2860	CGTCCGGTGTGACCGGGGGGCTCCCGAGAGCCGAGACATCCCATAGCAGGAGCACATGTGCTGT	2919
QY	3550	cccaaaagacactctgtgctccctgaaatccctgacagaggtcccccacaggtgcacagatgagaaggt	3609
Db	2920	CCCAAAAGACCTCTTGTCTCCCTGATATCTCTGTACAGAGGCGCCCCCAAGGAGCAATGTGAAGGG	2979
QY	3610	gtctgagagccccaagagccgggtgggtcctctgaaagaggtctccccaagaggtgtgtatcgtttcttc	3669
Db	2980	GCTGAGGCCCCAGGCGCGGGGGCTGTGGAAAGGCTCCCAAGATGTGTATCTGTCTCTC	3039
QY	3670	acgggcctctggtgtagctcccgagcgcccggtgaaccccaaggtctgtgacacacacccctgtgacac	3729
Db	3040	ACGGGCTCTGTAGTGAAGCGCCGACGCCCGGAGCCCAAGCTCTGTACACACACCCCTGTGACACC	3099
QY	3730	ccagagcaaacctgggtgggtgcacagccagagcgctcttcaaggtccgggtgcacagccggtgtggagag	3789
Db	3100	CCAGACAAACTGGGGGGGCAAGCAAGCAGGCGCGCTTCAAAAGTCTGGGAGCAAGCGGTGGGAGAG	3159
QY	3790	cccttaacccaaggtctgctctccagcctccagcaaaccccggtcgcgtctgtgtgccttcgac	3849

[illegible]

Db 4100 -----ACAGG 4104

Qy 4927 gccagtggacacaaagggccagtgaggaacaactctgtgtgagaggaagcaagcaaaag 4986

Db 4105 gccagtggacacaaagggccagtgaggaacaactctgtgtgagaggaagcaagcaaaag 4164

Qy 4987 atggagagctggggccctgcctccagagcccccagggcagggccctgcagagccagaca 5046

Db 4165 atggagagctggggccctgcctccagagcccccagggcagggccctgcagagccagaca 4224

Qy 5047 agggcacagaaacagcagggccacacaaactacagcagctattccaaagcgaagccctc 5106

Db 4225 agggcacagaaacagcagggccacacaaactacagcagctattccaaagcgaagccctc 4284

Qy 5107 atcggggcccggcgaagaaac 5166

Db 4285 actcgggcccggcgaagaaac 4344

Qy 5167 cgacaaacagagctgtgcctgcctgcctgcctgcctgcctgcctgcctgcctgcctgcctgc 5226

Db 4345 cgacaaacagagctgtgcctgcctgcctgcctgcctgcctgcctgcctgcctgcctgcctgc 4404

Qy 5227 tcccttgcagagcgtgaggtgcagacagcggcagcccgcccttctcacccttcgtgcgg 5286

Db 4405 tcccttgcagagcgtgaggtgcagacagcggcagcccgcccttctcacccttcgtgcgg 4464

Qy 5287 gtggagagcagagcgttccaccacacacacacacacacacacacacacacacacacacac 5346

Db 4465 gtggagagcagagcgttccaccacacacacacacacacacacacacacacacacacacac 4524

Qy 5347 ccaagccccacagaaagccttcctcctgcctgcctgcctgcctgcctgcctgcctgcctgc 5406

Db 4525 ccaagccccacagaaagccttcctcctgcctgcctgcctgcctgcctgcctgcctgcctgc 4584

Qy 5407 tcccttgcagagcgtgaggtgcagacagcggcagcccgcccttctcacccttcgtgcgg 5466

Db 4585 tcccttgcagagcgtgaggtgcagacagcggcagcccgcccttctcacccttcgtgcgg 4644

Qy 5467 cggcccttcctgccttcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 5526

Db 4645 cggcccttcctgccttcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 4704

Qy 5527 agtaccttccttccttccttccttccttccttccttccttccttccttccttccttccttcct 5586

Db 4705 agtaccttccttccttccttccttccttccttccttccttccttccttccttccttccttcct 4764

Qy 5587 gacctctgtggccctactacactgaacacactgcctcccaaaagaaagcaaacactcaag 5646

Db 4765 gacctctgtggccctactacactgaacacactgcctcccaaaagaaagcaaacactcaag 4824

Qy 5647 gagaaggtgcggccagaaagcactgtgagagggcctcgtccgcttgagagaaacactc 5706

Db 4825 gagaaggtgcggccagaaagcactgtgagagggcctcgtccgcttgagagaaacactc 4884

Qy 5707 aaaggtcccgagtgagcagctgcgcacactgcgcagagcccccagggcagggccagggc 5764

Db 4885 aaaggtcccgagtgagcagctgcgcacactgcgcagagcccccagggcagggccagggc 4944

Qy 5765 gctgagccggcgaagcagggccactgcgcacactgcgcagagggcctgcgcagggcgtg 5824

Db 4945 gctgagccggcgaagcagggccactgcgcacactgcgcagagggcctgcgcagggcgtg 5004

Qy 5825 cagagctgtactgtgtgagcgggagagatggggcgagagggagggcagccagccagc 5884

Db 5005 cagagctgtactgtgtgagcgggagagatggggcgagagggagggcagccagccagc 5064

Qy 5885 aaggtgccaacatgagtgagcagaaagggcctccgagagagcccgggggggcccgag 5944

Db 5065 aaggtgccaacatgagtgagcagaaagggcctccgagagagcccgggggggcccgag 5069

Qy 5945 gagcactgggtgcatgagggcctgtgcctgtggaccggcgggctactcctgtgtgaggg 6004

Db 5070 --ACACUGGGTGCATGAGGGCTGTGCGGTGTGGACCGGGCGGTCTACTGCTGGTGGCGGG 5127

QY 6005 aagctttttgggctgcag 6022

Db 5128 AAGCTTTTGGGCTGCAG 5145

RESULT 7

MUSMA

LOCUS MUSMA 7222 bp mRNA ROD 07-FEB-1999

DEFINITION Mouse mRNA for unknown product, complete cds.

ACCESSION D29801

VERSION D29801.1 GI:475015

KEYWORDS OFF.

SOURCE Mus musculus postnatal-day1 brain cDNA to mRNA, clone\_lib:library of Stratagene (La Jolla, CA).

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 7222)

AUTHORS Imai,Y.

TITLE Direct Submission

JOURNAL Submitted (04-APR-1994) to the DDBJ/EMBL/GenBank databases. Yuji Imai, Osaka University Medical School, Molecular Neurobiology, Tanabe; 2-2 Yamadaoka, Suita, Osaka 565, Japan (Tel:06-879-3646, Fax:06-879-3648)

REFERENCE 2 (bases 1 to 7222)

AUTHORS Imai,Y., Suzuki,Y., Matsui,T., Tohyama,M., Wanaka,A. and Takagi,T.

TITLE Cloning of a retinoic acid-induced gene, GRI, in the embryonal carcinoma cell line P19: neuron-specific expression in the mouse brain

JOURNAL Brain Res. Mol. Brain Res. 31 (1-2), 1-9 (1995)

MEDLINE 96078271

COMMENT Submitted (04-Apr-1994) to DDBJ by: Yuji Imai

Department of Molecular Neurobiology, Tanabe

Osaka University Medical School

2-2 Yamadaoka, Suita

Osaka 565

Japan

Phone: 06-879-3646

Fax: 06-879-3648.

FEATURES

source

1..7222

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone\_lib="library of Stratagene (La Jolla, CA)"

/dev\_stage="postnatal-day1"

/tissue\_type="brain"

309..5831

/codon\_start=1

/product="Unknown"

/protein\_id="BAA06184.1"

/db\_xref="GI:475015"

/translation="MQSFRRCRGGHKKQNPQTSQETSRLENYRPGAGLSCDROR

LLADYVSPQVTGEGGTGPGSCVATAADKYHRSKSLQGRPAFPSPYVSDSPY

GRYSGEGLTWGEGGPPPPQPLPGANSKYENMLKTYVPPNRPQYEGQPLPF

RTHSLHVPFPQPPQPLKLPORPNDLSPLEFPQGHFPQHSQSFSTSTTAPT

VGGGQGHAYSKYCTAPSAQPHDPSANLANPAGVQNLHAYQGRGLYEQQAL

QCRHHTQTHLYONLAKYHQGGQGYCPDPTAVRTPEQYTYTTFSSSHSPARSV

RSPSYSTPSPPLNLENFPYSSQPLSTGAPPTGTDHSHFPLNLPSPDAAASVD

CAGNCKPLOKEKLPDLNLSLVLSLQSLTALTSOVENTSNVQVLLSKATMPQKGVN

LYSTRPQHSQCHSPSGYSAPAGTPLSEPPSPSTQSTHAPQDQTYLSSGSDPL

ERSFLYCSQARGSPARVNSKAPESVSTCSPTPDDMSTKSDDSFSLHSTLPDS

TSKFLYQERDCPRLLSLAQEDLSEILGLQEAIVERKADKANAEASSLPKNDGKPP

SLNHHGACLTVAKTWSQGPETPLPELQDKGSKDFSGPLFEDFSVAFTDP

KTSSTSPGTGTLGTATPDTTAAFDCTPDTTASVDGKGFANFAMPEENLGDAPR

WGLHPGELTGLEQKASDGVKADAEASACMGFOEDHAIKGAALSGDFKQEA

EGVKEFVGLLQCPYVAKADQWLEESRHCSTSDGDLPLLPGRKEDLEAEYSS

LCLELGFEPQSLDPLSPKAPLACMTKEAEADLTAGWVSPCHLSGEPALVIGPS

VGASQVQSFPESFSSHKPGEPEPNERAPSSSTSGSLAPKPKPAVPEGTAKK

EVFPGKSLSRVHRGLPEAEADSPFPAIPALPDLPLPESCTGPPQGAEGAGPGR

LSEGLPMTCTRSITALSFPOTGPPGLTTPPTPDKLGKQKRAAFKSGKRVGKPSKA

AASSPSPNPAALPVADSSPMGSKTKPEPSPSPKQDSWLSRSTKPPQVFEHAKRRP

CDS



QY	2326	gggctgaggaagccatcgtgagaagcgacaaagcttgggtgaagcaccagcgtg	2385	gcccgaagcccaacaagcgtctgtcccgagcgcccatcgcacaaagaagcgtctgtg	3465
QY	2327	gggctgaggaagccatcgtgagaagcgacaaagcttgggtgaagcaccagcgtg	2386	gcccgaagcccaacaagcgtctgtcccgagcgcccatcgcacaaagaagcgtctgtg	3466
DB	2094	ggactgcaggaagaccattgttgagagagctgacaaagccgtgggctgaggcttccagccgtg	2153	gcccgaagcccaacaagcgtctgtcccgagcgcccatcgcacaaagaagcgtctgtg	3174
QY	2386	gtcaagacagcagcaagcaacccctctcgtgtggaacacacagcgcgtcgtgactct	2445	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3525
DB	2154	cccaaggacaaatggcagacaccccttctacatggagaacacatggcggcctctggagact	2213	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3526
QY	2446	gtggccaagagtggtgtgcccggcgtggtggagccgagcctgcccactctcttgag	2505	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3527
DB	2214	gtagcaaaagactttatggtcacagccaggagagccagacacccctccctgagccttgag	2273	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3528
QY	2506	ctggacaagggcggaatgccaagagacttaagcccgagcgtgttttgaagacccttcgtg	2565	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3529
DB	2274	ctggacaaggggtggcagacacacagagactttcagcccgggcgtgtttggaagacccttctg	2333	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3530
QY	2566	gctctgtacgctgaccccaaaaagaactgtctctctctcttggatcacaagccc	2625	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3531
DB	2334	gcttggccacactgaccggagaaacacagcagctccctgtcttccggcaccagccct	2393	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3532
QY	2626	accttgggggttctgtccagacccccactacagcagcttttgaactgttccggagaca	2685	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3533
DB	2394	tttgtagggactggccactctctgacccacacagcagcatttgaactttccagataca	2453	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3534
QY	2686	acgctggcagctcagcgagcagcccaaccccttgcctggcagaggaacactgggg	2745	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3535
DB	2454	cccaactggcagctcggttgatgtgcccaccccttggcctggccagagagacactgggc	2513	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3536
QY	2746	gatgcttcccaggtgggaattgcaacctggcgagcttaccagaggcctggagcaaggt	2805	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3537
DB	2514	gatgcttggcctgtgtggggcctccacacccctgctgagcttaccagaggcctgggaacaggt	2573	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3538
QY	2806	gggaagcctcagatggatcagcaaaagggacacccatgagcttggcgtcctgggc	2865	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3539
DB	2574	gcuaaaagccttggcagcggcgtggcgaagcagcagatgcacagagagcctctgcttgcactggcc	2633	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3540
QY	2866	ttccagagaggaacccctcgtggggaagctggcctcgttgcggggagacttcaagcag	2925	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3541
DB	2634	ttccagagaggaacccctcgtggggaagctggcctcgttgcggggagacttcaagcag	2926	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3542
QY	2926	gagagagtggttgggtgaagagagagcaggtgggtgctgcaagtggccagagtgccc	2985	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3543
DB	2694	caggagcagagagggggtgaagagagagtggtgggtgctgcaagtggccagagtgccc	2753	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3544
QY	2986	aagctgaacgctggctggagagcagcggcactgctgttccaccccgagacttggggac	3045	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3545
DB	2754	aaggccacacactggctggagagaccgcacattgctgttccctcactgacttccggagac	2813	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3546
QY	3046	ctcccaactgctgcaccccaacaggaagagagcactggaagctgagagaggtactcc	3105	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3547
DB	2814	ctggcactgtgcccgccttggcaggaagagagacacttggagagctgaagagaggtactct	2873	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3548
QY	3106	tcctatgtgactcctggcagcccacagcagagcgtggtgcatcaggacccgctgtca	3165	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3549
DB	2874	tctctgtgtgaactgctgggagcccttgagcagagagccacagcctgcagagcccggttctcc	2933	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3550
QY	3166	cccaagcccacactcatcgcacaaagggaggttggagaggtgctggactccaagggcc	3225	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3551
DB	2934	cccaagcccacactcatcgcacaaagggaggttggagaggtgctggactccaagggcc	2993	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3552
QY	3226	ggctgggtctcctgtgcaacctcaggggagtcctgtactcgtgggcctacagtg	3285	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3553
DB	2994	ggctgggtctcctgtgcaacctcaggggagtcctgtactcgtgggcctacagtg	3053	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3554
QY	3286	ggcaccaggtcaaggtccagcgtgttggactcctctgtcacaatgaagcaggt	3345	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3555
DB	3054	ggtggcccaatcaaggttcagagtggttgaatttcttctgttcacacatgaagccagga	3113	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3556
QY	3346	gaagaggggctgatggggagagctccaggggattccacacactcctggagcctctctg	3405	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3557
DB	3114	gaagaggggctgatggggagagctccaggggattccacacactcctggagcctctctg	3173	ccacggggcaaaagcttacgagcgtctcggttgccacgggggtctgccagagccgagac	3558



Db	771	CCCCAGGCTGACGGCCAGCTGACCCGGCCAGACAGGGCCCATCTGCGACCGCTGGCC	780
QY	5805	ggggcctgtccggaggctgcagagctgtactgtgtatgcccgggaggtatgggggagc	5864
Db	781	GGGGCTGTCCGGAGGCTCGAGAGTGCTACTGCTGTATCGCCGGGAGATGGGGCG	840
QY	5865	aggaggcagcccccagccgacaaagggctcgcaaacatagtcgcagcaagaggctccggcag	5924
Db	841	AGGAGGACGCCCCAGCCGACAAAGGGTCGCAACATGATGTCAGCAAGAGGCTCCGGCAG	900
QY	5925	agcccgccggggaggcccaaggaaactgggtgcataagacctgtgcgtgtgacacggcg	5984
Db	901	AGCCCGCGGGGAGGCCCGAGAGCACTGGGTGCATGAGCGCTGCCGTGTGGACCGCG	960
QY	5985	gcgtcacctgdtggccgggaagctctttgggctgcag	6022
Db	961	CGCTCTACCTGTGTGGCCGGGAAGCTCTTTGGCTTGCAG	998

RESULT	9				
AF021115/c					
AF021115					
LOCUS					
			595 bp	DNA	STS
					03-APR-1998

VERIFICATION AF021115  
 ACCESSION AF021115  
 VERSION AF021115.1  
 KEYWORDS STS.

ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostom	
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.	

**AUTHORS**  
Philibert, R. A., Horelli-Kuitunen, N., Robb, A. S., Lee, Y. H.,  
Long, R. T., Damschroder-Williams, P., Martin, B. M., Brennan, M. B.,  
Parotie, A. and Ginns, E. I.

containing genomic cosmid clones  
JOURNAL Eur. J. Hum. Genet. 6 (1), 89-94 (1998)  
MEDLINE 98454304

**AUTHORS**  
 Philibert, R.A., Hörelli-Kuitunen, N., Robb, A.S., Lee, W.Y.-H.,  
 Long, R.T., Damschroder-Williams, P., Martin, B.M., Brennan, M.B.,  
 Palotie, A. and Ginns, E.I.

JOURNAL Submitted (28-AUG-1997) Clinical Neuroscience Branch, National Institutes of Mental Health, 49 Convent Drive Rm B1E16, Bethesda MD 20892, USA

```
source
1..595
/organism="Homo sapiens"
/db_xref="taxon:9606"
```

```

/map="17q11.2"
/note="primers to amplify repeat: forward:
ACAGACCGACTGCCAGCCCAT; reverse:

```

BASE COUNT	90 a	140 c	230 g	120 t	15 others
ORIGIN					

Query Match 8.2%; Score 495; DB 11; Length 595;  
Best Local Similarity 93.6%; Pred. No. 2.3e-77;  
Matches 557; Conservative 11; Mismatches 18; Indels 9; Gaps

**OY** 1019 ggcgccccagcagccctg-gcatatcccgaagctccaaggcagaagctgcagaacgac 107  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
**Dd** 595 CACCGCCCCAGCAGCCCTGNGCATACCCCAGAAGCTCCAAGGCAGAAGCTGCAGAACGAC 536

1078 attgctccctctgccttccccaggraccacatttctcagcattcccagtccttc 1137  
|||||  
535 ATTGCTCCCTCTGCCTTCCCCAGGGTACCCACTTTCCTCAGCATTCCTCAGTCTTC 476

QY 1138 ccacctcctccacctactcctcctctgtccagggtggtagggcagggggccacctcctat 1139



6.5%; Score 391.8; DB 11; Length 495;

Best Local Similarity 94.2%; Pred. No. 3.9e-59;  
Matches 469; Conservative 1; Mismatches 21; Indels 7; Gaps 6;

denature: 30 seconds at 94 degrees C  
annealing: 75 seconds at 55 degrees C





[illegible]



```
repeat_region /note="AluY repeat: matches 1. .291 of consensus"  
15726. .15823  
/note="49 copies 2 mer aa 65 conserved"  
repeat_region /note="L1M6 repeat: matches 5759. .5805 of consensus"  
15884. .15931  
repeat_region /note="L1M6 repeat: matches 1. .301 of consensus"  
16192. .16486  
repeat_region /note="AluY8 repeat: matches 1. .310 of consensus"  
16525. .16834  
misc_feature /note="AluY8 repeat: matches 1. .310 of consensus"  
complement(1676. .17109)  
/note="match: GSS: Em:A0696298"  
17613. .17930  
repeat_region /note="AluX repeat: matches 1. .310 of consensus"  
18031. .18170  
repeat_region /note="L1M4 repeat: matches 4457. .4612 of consensus"  
18193. .18477  
repeat_region /note="AluX repeat: matches 1. .291 of consensus"  
18483. .18799  
repeat_region /note="L1M4 repeat: matches 7630. .7972 of consensus"  
18832. .18938  
repeat_region /note="L1M4 repeat: matches 7445. .7560 of consensus"  
19080. .19145  
repeat_region /note="MER5A repeat: matches 44. .112 of consensus"  
19226. .19516  
repeat_region /note="L1 repeat: matches 1921. .2236 of consensus"  
19543. .19716  
repeat_region /note="MIR repeat: matches 79. .256 of consensus"  
19795. .20002  
repeat_region /note="L1 repeat: matches 1618. .1833 of consensus"  
20233. .20387  
repeat_region /note="L1 repeat: matches 1456. .1583 of consensus"  
20632. .20683  
repeat_region /note="L1 repeat: matches 2677. .2729 of consensus"  
21378. .22070  
repeat_region /note="L1 repeat: matches 2. .741 of consensus"  
22071. .22382  
repeat_region /note="AluX repeat: matches 1. .309 of consensus"  
22383. .22393  
repeat_region /note="L1 repeat: matches 741. .751 of consensus"  
22394. .22673  
repeat_region /note="AluX repeat: matches 1. .282 of consensus"  
22674. .22840  
repeat_region /note="L1 repeat: matches 751. .928 of consensus"  
22841. .23138  
repeat_region /note="AluY repeat: matches 1. .297 of consensus"  
23139. .23585  
repeat_region /note="L1 repeat: matches 928. .1231 of consensus"  
23738. .24015  
repeat_region /note="AluY8 repeat: matches 1. .313 of consensus"  
24067. .24116
```

Query Match 1.88; Score 110; DB 9; Length 43761;  
Best Local Similarity 52.78; Pred. No. 4.2e-10;  
Matches 295; Conservative 0; Mismatches 250; Indels 15; Gaps 2;

```
Db 5284 ATGCAACTGAGGAGGAGATCTCCAGTGTGGGCAAGCAGTGTCCATGAGGTTCCAGAAAC 5225  
QY 1639 gcagygatcaatgacacacagcaatcattccctgtctaatccctcccaagcagatgc 1698  
Db 5224 AGATTTTA-----CAGTTAATGCTCACTCACTCACTCACTCACTCACTCACTCACTCA 5177  
QY 1699 accagctctgtgagaccacagcagcgtgagcagcagcagcagcagcagcagcagcagcagc 1758  
Db 5176 CCGAGTCT---AATTCTATGCTCAGGCTTCAAGGAGTTGAGCTAGAGGAGGATCA 5120  
QY 1753 gagaactctgtcgtatctcagcctgcagagcctcagcagcagcagcagcagcagcagcagc 1818  
Db 5119 GAAAGCAGCTGAGAGATCTCTGCTTGTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5060  
QY 1819 aacatccacacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1878  
Db 5059 AATCTTCTAAGACTGCTCAGCAGCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5000  
QY 1879 ggtcagaagacccctgtgtc 1898  
Db 4999 ACCTCCAGAGGAGGCTTCAATC 4980
```

Search completed: December 4, 2001, 16:20:21  
Job time: 11264 sec

```
QY 1339 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1398  
Db 5524 CAACAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5465  
QY 1339 cattaacaaacacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1458  
Db 5464 CAGTATATCTAGCTGACCCAGAGCTGCTGCAAAAGCAGATGGGCGATGACACACAG 5405  
QY 1459 ccgagacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1518  
Db 5404 CCTGAGGTCTCTGAGGTCTCTGAGGTCTCTGAGGTCTCTGAGGTCTCTGAGGTCTCTGAG 5345  
QY 1519 cactacacccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1578  
Db 5344 CCTTCCAGAGTCTCTGAGGTCTCTGAGGTCTCTGAGGTCTCTGAGGTCTCTGAGGTCTCTGAG 5285  
QY 1579 atgcacaaactgtgaaacttccctcagcagcagcagcagcagcagcagcagcagcagcagc 1638
```